The relationship between maternal diet and gestational diabetes mellitus

A metabolomic approach

Jamie Violet de Seymour

A thesis submitted in complete fulfilment of the requirements for the degree of Doctor of Philosophy in Biomedical Science, The University of Auckland, New Zealand, 2017.

Abstract

Gestational diabetes mellitus (GDM) is a carbohydrate intolerance first recognised during pregnancy and is associated with serious health consequences to both mother and offspring. The best approach to reduce the adverse consequences of GDM is to prevent its development. Despite evidence suggesting a role of maternal diet in GDM development, there is little agreement on the best approach to prevent its occurrence. The observational nature of dietary research has made it difficult to determine mechanisms underlying the associations. Metabolomics, the study of low-molecular weight compounds, is an analytical approach that could be used to explore how maternal diet is associated with GDM.

Aim: The aim of this thesis was to investigate the relationship between maternal diet and GDM, by exploring the effect of both diet and disease on the maternal metabolome, in a Southeast Asian multi-ethnic population (N=1,247) from Singapore (The GUSTO cohort).

Method: Maternal dietary pattern analysis was conducted from 24-hr dietary recalls, to explore associations between maternal diet and GDM. The effect of GDM on metabolism was determined by metabolite profiling of maternal plasma and hair, using gas chromatography-mass spectrometry. The dietary pattern data and metabolomic data were combined to look for associations between maternal diet and GDM.

A seafood-noodle-based dietary pattern, consisting of higher amounts of soup, fish and seafood products, and noodles; and lower amounts of white rice, curry-based gravy, legumes/pulses, and ethnic bread, was found to be significantly associated with a reduced likelihood of GDM ($P < 0.01$). Metabolite profiling of maternal hair revealed 277 metabolites, none of which were associated with GDM. Of the 219 metabolites detected in maternal plasma, 26 were associated with GDM. One metabolite, 2-hydroxybutyric acid, was significantly higher in GDM cases ($P < 0.001$, FDR = 7.04E-07) and could discriminate GDM outcome with higher accuracy than the model of clinical risk factors. Metabolic pathway analysis showed that a combination of the significant GDM metabolites may be linked to dysfunction in mitochondrial oxidative phosphorylation. Two plasma metabolites had a significant linear association with the seafood-noodle-based dietary pattern. However, this significance was removed after adjustment for multiple comparisons. Two significant canonical roots were derived from canonical correlation of metabolites and food groups but explained only 1% of the variation in the food groups and 4% of the variation in the metabolites, deeming them unlikely candidates as dietary biomarkers of GDM development.

Conclusion: A seafood-noodle-based dietary pattern was associated with GDM in this multi-ethnic Asian cohort. The pattern was distinct from those found previously, emphasising the importance of applying caution when extrapolating dietary findings outside of the studied population. The plasma metabolome could differentiate women that developed GDM from those that did not, linking GDM to a potential dysregulation in mitochondrial function. Although the diet-metabolite associations were not strong enough to deem them clinically useful as dietary biomarkers of GDM, they highlight some interesting associations. Future work should validate these findings in early pregnancy.

Acknowledgements

Firstly, I would like to acknowledge the support and guidance of my supervisory team: Professor Philip Baker, Dr Cathryn Conlon, Associate Professor Clare Wall, and Dr Elizabeth McKenzie, and my advisory team; Associate Professor Silas Villas-Bôas and Dr Karolina Sulek. Without your expertise, high expectations, and continued encouragement, I would not have made it through these past few years. Thank you for believing in me, and providing me with opportunities to grow and develop as a student, a scientist, and an academic.

I would also like to gratefully acknowledge the Agnes Paykel Trust for financially supporting my doctoral studies, the Ministry of Business, Innovation, and Employment for funding the research conducted in this thesis, and both the Faculty of Medical and Health Science – Postgraduate Student Association and the Food for Health Programme for financial support to attend and present at international conferences. I would also like to acknowledge the financial assistance that I received from the Prime Minister's Scholarship for Asia which allowed me to conduct research in Singapore, and form strong international collaborations.

To our international collaborators from the National University of Singapore (Dr Mary Chong, Dr Gerard Wong, and Dr Neerja Karnani), thank you for sharing your time and knowledge throughout this process, and welcoming me to Singapore. To the families in the GUSTO study, thank you for your participation. Without your contributions to science, this work would have not been possible.

To the students that helped to make the experimental mountain conquerable; Airu, Aude-Claire, Thibaut, and Jena, I cannot thank you enough. I would also like to acknowledge my third-party proof-reader, Hilary van Uden, and offer my sincere thanks to my biostatistician Dr Beatrix Jones for providing your expert guidance and support for the statistical components of my research, it has been a pleasure learning from you.

My most heartfelt gratitude is reserved for my close friends and family. For your continued patience, perseverance, and commitment to our friendship throughout the most difficult few years. For accepting and supporting me through each struggle, and celebrating each success. A special mention to Kim, Missy, and Lydia – your friendship means the world to me.

To Mum and Dad (+ Nick and Leanne), thank you for helping shape me into the independent, driven, (sometimes stubborn), lady that I am today. You have always encouraged me to be the best I can be, and I hope that I have made you proud.

To my siblings - follow your passions and chase your dreams. I know each of you will achieve greatness.

Last, but certainly not least, to Ian (and Dexter [doggo]) for your unconditional love and support through the most daunting of times, for keeping me accountable, and reminding me of my strength and worth when I most needed it. You warm my heart. Looking forward to the next chapter *(pun intended).*

Table of Contents

List of Tables

List of Figures

List of Abbreviations

25(OH)D	25-hydroxyvitamin D
4-HPA	4-hydroxyphenylacetate
AMDIS	Automated deconvolution and identification system
ATP	Adenosine triphosphate
AUC	Area under the curve
BCAA	Branched chain amino acid
BMI	Body mass index
CDF	Common data format
CI	Confidence interval
CVD	Cardiovascular disease
DASH	Dietary approaches to stop hypertension
EDTA	Ethylenediaminetetraacetic acid
EFA	Exploratory factor analysis
FDR	False discovery rate
FFQ	Food frequency questionnaire
FIA-MS/MS	Flow injection analysis–tandem mass spectrometry
FIE-MS	Flow injection electrospray–mass spectrometry
GC	Gas chromatography
GC-MS	Gas chromatography-mass spectrometry
GDM	Gestational diabetes mellitus
GI	Glycaemic index
GL	Glycaemic load
GUSTO	Growing up in Singapore towards healthy outcomes
HMBOA	2-hydroxy-7-methoxy-1,4(2H)-benzoxazin-3-one
HUMBA	Healthy mums and babies
IDF	International Diabetes Federation

LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
MCCV	Monte Carlo cross validation
MCF	Methyl chloroformate
MS	Mass spectrometry
MSTFA	N-methyl-N-(trimethylsilyl)trifluoroacetamide
N	Number
NAD	Nicotinamide adenine dinucleotide
NIPPER	Nutritional intervention preconception and during pregnancy to maintain healthy glucose levels and offspring health
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance
NUS	National University of Singapore
OGTT	Oral glucose tolerance test
OR	Odds ratio
PCA	Principal component analysis
QC	Quality control
RCT	Randomised controlled trial
ROC	Receiver operating characteristic
ROS	Reactive oxygen species
SD	Standard deviation
SVM	Support vector machine
TCA	Tricarboxylic acid
TMAO	Trimethylamine N-oxide
UPLC-MS	Ultra performance liquid chromatography – mass spectrometry
WHO	World Health Organisation

Chapter One: Introduction and Literature Review

1.1 Introduction

Gestational diabetes mellitus (GDM) is a pregnancy disorder that is associated with many adverse short- and long-term consequences for both the mother and her offspring. Typically diagnosed later in pregnancy (around the 24th - 28th weeks of gestation), the metabolic changes associated with GDM development have already occurred, and the fetus will have been exposed to the effects of maternal hyperglycaemia. As a result, the consequences of GDM may have already taken effect before GDM is diagnosed and treatment sought. The most commonly used screening procedure in early pregnancy to identify women as at-risk of GDM development is the assessment of clinical risk factors, which include high body mass index (BMI), advanced age, previous history of GDM, and/or a family history of diabetes. Research has shown that these clinical risk factors are inadequate for predicting almost half of the women who will go on to develop GDM (Chong et al., 2014). To date, no other screening tools have been successfully used in the clinical setting to identify women at risk of GDM development later in pregnancy. The prevention of GDM cannot be accomplished without early identification of women who would benefit from intervention. A blanket approach to prevention would not be cost effective for the public health system, and may risk adverse outcomes by intervening in pregnancies that were not at risk of GDM. To reduce the associated consequences of GDM, prevention would be the optimal strategy. Therefore, early effective screening for GDM should be investigated in parallel to prevention strategies, to determine which women should be targeted for preventative interventions to prevent GDM development and its associated consequences.

An effective method for the prevention of GDM has also not yet been established. Research has demonstrated a profound effect of diet on metabolism, type 2 diabetes risk and management, and therefore offers a promising avenue to explore for GDM prevention (American Diabetes Association et al., 2008; Hu et al., 2001; Ajala et al., 2013). There is an abundance of literature demonstrating an association between the maternal diet and GDM (Table 1.3). However, the results of these studies together are inconclusive, and, in several cases, conflicting. The majority of the research has been conducted in an observational manner, using subjective, self-reported dietary assessment measures. In addition, many studies have focused on a single aspect of the diet, considering nutrients as individual and independent entities, which is not the case in a typical human diet.

Dietary pattern analysis is an alternative approach to analysing dietary data; this method takes into account the complete set of foods consumed, and therefore respects how the nutrients in foods interact with one another. Recent research into dietary patterns and GDM has provided some valuable insight in the field (Table 1.4) and warrants further investigation in different populations.

Metabolomics is the study of the low molecular-weight compounds present in a biological system. It is an analytical tool that could prove valuable in the investigation of the association between the diet and GDM, and can be used to identify potential biomarkers of GDM. Metabolomics has been used successfully to differentiate women who developed pregnancy disorders, including pre-eclampsia and fetal growth restriction, from those who did not, thus demonstrating its potential as a screening tool for pregnancy disorders (Horgan et al., 2011; Kenny et al., 2010). Only a small number of studies have explored its capabilities in GDM, with mixed results (Table 1.6). These studies have proven successful for differentiating women who developed GDM from those who did not, however, the metabolites identified have differed between studies and no consistent clinical biomarkers have been identified. There are two possible explanations for these discordant findings. Firstly, the majority of the studies conducted thus far have been performed on small sample sizes, which makes results highly susceptible to random effects. Secondly, there is a lack of standardisation within metabolomics research, and findings were generated on different analytical platforms, with different sample processing methods, and different statistical analyses. Future research searching for biomarkers of GDM development requires investigations using large study groups and well-planned, replicable experimental designs.

The metabolism and, in turn, the human metabolome, is highly influenced by dietary intake. Therefore, the metabolome is a logical place to investigate associations between maternal diet and GDM development, by examining associations between diet and the metabolome. Using the metabolome to link the associations between maternal diet and GDM could assist with understanding *how* maternal diet is related to the metabolic dysregulation associated with GDM. Understanding the link between the diet and GDM, through the metabolism, will highlight metabolic pathways that could be targeted with preventative dietary interventions.

1.2 Gestational Diabetes Mellitus

1.2.1 Definition and pathogenesis.

By definition, GDM is a carbohydrate intolerance that is first recognised during pregnancy (National Diabetes Data Group, 1979). Carbohydrate intolerance is an impaired ability to respond appropriately to glucose load. There are two accepted pathological routes of carbohydrate intolerance in GDM: insulin resistance, which is the inability of the body's cells to respond to insulin secreted from the pancreas; and impairment of pancreatic beta cell response to glucose load, resulting in insufficient production of insulin to appropriately metabolise the available glucose (Metzger et al., 2007). The pancreas releases insulin in response to an increase in blood glucose levels, and the dysregulation that can occur in GDM can be either an insufficient response to the glucose, or a reduction in pancreatic beta cell mass, resulting in an inability to secrete the levels of insulin required. The result of carbohydrate intolerance is a high maternal blood glucose level – hyperglycaemia. Unlike other forms of diabetes, GDM is specific to pregnancy and tends to resolve after delivery, although the mothers' risk of diabetes later in life is significantly increased following GDM. The exact mechanisms underlying the progression to GDM are still under investigation, particularly in regard to women who enter pregnancy without clinical risk factors. There is a clear association, however, between excess maternal adipose tissue and the development of insulin resistance and diabetes (Chu et al., 2007).

1.2.1.1 Insulin resistance in healthy pregnancy.

As pregnancy progresses naturally, many hormonal and metabolic changes occur. One such change is insulin sensitivity. Insulin sensitivity is increased in the early stages of pregnancy for the maternal tissues to uptake, utilise, and store glucose, in order to fuel the body for the increased needs of pregnancy, and to develop fat stores that are required later in gestation (Catalano, Ishizuka, & Friedman, 1998). As pregnancy progresses, the requirements of the fetus shift to high-velocity growth and the mother's insulin sensitivity reduces, resulting in a mild state of insulin resistance. This resistance is biologically and evolutionarily beneficial as it keeps the maternal blood glucose higher than normal so that glucose can be transported via the placenta to the fetus for utilisation. A mild state of maternal insulin resistance also means that the mother's fat stores are able to be broken down and used to fuel the fetus' growth. The insulin resistance that occurs as part of a healthy, uncomplicated pregnancy is thought to be responsible for GDM in individuals who are unable to adapt to reduced insulin sensitivity, or

that already have undiagnosed diabetes, pre-diabetes, or mild insulin resistance entering pregnancy (Mastrogiannis, Spiliopoulos, Mulla, & Homko, 2009). However, for those without an underlying risk entering into pregnancy, it is less clear what biological pathways lead to GDM development.

1.2.1.2 Insulin resistance and adiposity.

Obesity is the accumulation of excess adipose tissue. The development of obesity is associated with many adverse health complications such as heart disease, sleep apnoea, hypertension, osteoarthritis, and some cancers (Bray, 2004). There are strong links between excess adipose tissue and the development of insulin resistance and diabetes (Boyko, Fujimoto, Leonetti, & Newell-Morris, 2000; Neeland et al., 2012; Rana, Li, Manson, & Hu, 2007; Y. Wang, Rimm, Stampfer, Willett, & Hu, 2005). Adipose tissue acts as an endocrine organ, as it is capable of the production and secretion of a variety of hormones and adipokines (Kershaw & Flier, 2004), some of which have been directly related to the development of insulin resistance and diabetes. These include tumour necrosis factor-alpha, adiponectin, interleukin-6, resistin, and leptin (Freitas Lima et al., 2015).

Meta-inflammation (a type of chronic, low-grade metabolic inflammation; Hotamisligil, 2006) is associated with obesity has also been closely linked to the development of insulin resistance (Duncan et al., 2003; Luft et al., 2013). It is important to note, however, that not all adipose tissue is the same. Individuals with high amounts of subcutaneous adipose tissue are not at the same risk of insulin resistance and diabetes as individuals with high levels of intra-organ and visceral adipose tissue (Hanley et al., 2009; Miyazaki et al., 2002; Seidell, Björntorp, Sjöström, Kvist, & Sannerstedt, 1990). Studies have shown that visceral fat generates higher levels of interleukin 6, and therefore may play an important role in obesity-related meta-inflammation (Fontana, Eagon, Trujillo, Scherer, & Klein, 2007). The difference in risk associated with the compartmentalisation of fat stores may indicate a potential compensatory mechanism for increased fuel consumption in individuals that are able to store excess adipose tissue as subcutaneous adipose tissue rather than visceral or intra-organ fat. In addition, white and brown adipose tissue have distinct metabolic differences. Brown adipose tissue plays a role in thermogenesis and as such contributes to energy expenditure, and may pose a much smaller detrimental effect on insulin sensitivity and diabetes risk than white adipose tissue. In a study of 2,934 participants, brown adipose tissue was found in higher amounts in women, younger participants, and leaner participants (Lee, Greenfield, Ho, &

Fulham, 2010), and higher amounts of brown adipose tissue were associated with lower levels of fasting plasma glucose than high amounts of white adipose tissue.

1.2.1.3 Mitochondrial dysfunction and diabetes.

Mitochondria are the powerhouses of the cell, responsible for many processes concerned with energy metabolism. They are thought to play a role in the development and progression of type 2 diabetes and GDM, metabolic diseases where energy metabolism is dramatically affected. Studies into the relationship between mitochondria and diabetes have found that participants with type 2 diabetes have reduced tricarboxylic acid cycle activity and impaired oxidative phosphorylation in skeletal muscle mitochondria, as well as an overall reduction in the size and number of mitochondria present (Højlund et al., 2003; Kelley, He, Menshikova, & Ritov, 2002; Ritov et al., 2005). Interestingly, Kelley et al. (2002) found similar mitochondrial dysfunction in obese participants without diagnosed diabetes, supporting the idea that mitochondrial dysfunction may be occurring prior to the development of diabetes. Additional support for the role of mitochondrial dysfunction in the development of diabetes comes from the work of Sparks et al. (2005) who found that a high-fat diet in insulin-sensitive humans administered acutely, significantly down-regulated a number of enzymes involved in mitochondrial oxidative phosphorylation. Similar results were obtained when a high fat diet was administered to C57Bl/6J mice for three weeks (Sparks et al., 2005). These findings offer valuable links between diet, metabolic dysfunction, and progression to diabetes.

1.2.2 GDM prevalence and risk factors.

The worldwide prevalence of GDM is difficult to quantify, as diagnostic criteria are not standardised and many countries do not practise universal screening. Estimates from the International Diabetes Federation (IDF) have reported that 1 in 7 births are affected by GDM – equivalent to 20.9 million children annually exposed to GDM and its associated consequences (IDF, 2015). South-East Asian women have some of the highest rates of GDM, estimates are as high as 24% of all live births are exposed to maternal hyperglycaemia during pregnancy (IDF, 2015). It has been suggested that the reasons for increased prevalence in this group is related to changes in nutrition as a result of a dietary shift from traditional food consumption to incorporation of Western foods and food habits, along with reduction in physical activity patterns as a result of improvements in technology and reductions in manual labour (Hu, 2011). Asian women and men are more susceptible to adverse health

consequences arising from increase in BMI at much lower increments than Caucasians (Deurenberg, Deurenberg-Yap, & Guricci, 2002; Deurenberg-Yap, Schmidt, van Staveren, & Deurenberg, 2000; Shai et al., 2006). Studies have shown that Asian women tend to carry higher amounts of visceral fat compared with Caucasians, which could explain the increased risk of metabolic complications such as GDM (Lim et al., 2011; Y.-W. Park, Allison, Heymsfield, & Gallagher, 2001; Rush et al., 2007).

In addition to ethnicity, other clinical risk factors for GDM development include advanced age, maternal BMI, previous history of GDM, and a family history of diabetes. Studies have found that maternal age ≥ 25 years can be considered a risk factor for GDM and ≥ 20 years in South Asians (Lao, Ho, Chan, & Leung, 2006; Makgoba, Savvidou, & Steer, 2012). Advancing age as a risk factor of GDM is not surprising given that ageing over the life course is associated with an increase in insulin resistance (Barbieri, 2001), and therefore mothers of advanced age are likely to be entering into pregnancy with a higher level of insulin resistance than their younger counterparts.

As discussed previously, adiposity has been found to be directly associated with diabetes and insulin resistance, and a high maternal BMI has been found to be a significant risk factor for the development of GDM (Chu et al., 2007; Sebire et al., 2001). More recently it has also been recognised that excess maternal weight gain during pregnancy, particularly in the first trimester, increases the risk of GDM development (Hedderson, Gunderson, & Ferrara, 2010). Weight gain is very closely related to energy balance and dietary intake, as is adiposity, and many studies have investigated the association between the maternal diet and GDM.

Despite strong associations between the recognised clinical risk factors and GDM development, women who do not fit into any of these high-risk categories are still developing GDM. A study conducted in Singapore found that high-risk clinical indicators failed to detect almost half of the women who later developed GDM (Chong et al., 2014). Without appropriate identification of at-risk mothers, it is difficult to deliver interventions to the women who require it the most. A blanket approach to prevention would be very costly to public healthcare systems, and could run the risk of inappropriate interventions in otherwise healthy pregnancies.

1.2.3 Consequences.

The development of GDM is associated with adverse consequences both to the mother and her offspring, in the short- and long-term, as well as to the public health system, and overall

societal functioning. Dysregulated carbohydrate metabolism in GDM results in an excess of circulating maternal blood glucose. High blood glucose levels – hyperglycaemia – can be detrimental to the health of the mother and unborn child. Table 1.1 lists the major adverse consequences of GDM development. The most marked adverse effect on the mother is a significantly increased risk of developing type 2 diabetes following delivery, with a relative risk as high as seven times that of women with normoglycaemic pregnancies (Bellamy, Casas, Hingorani, & Williams 2009). In addition to the risk of metabolic complications, a relationship has also been observed between GDM and increased likelihood of postnatal depression (M. Silverman et al., 2017).

Table 1.1

The Adverse Consequences Associated with the Development of Gestational Diabetes Mellitus

Short-Term	**Long-Term**
Macrosomia and birth complications	Infant obesity
Newborn respiratory distress syndrome	Cardiovascular disease (CVD)
Preterm birth	Metabolic syndrome
Newborn hypoglycaemia	Type 2 diabetes
	Cognitive impairment of offspring

(References: Bellamy et al., 2009; Boney, Verma, Tucker, & Vohr, 2005; Casey, Lucas, Mcintire, & Leveno, 1997; Hedderson, Ferrara, & Sacks 2003; Langer, Yogev, Most, & Xenakis, 2005; Robert, Neff, Hubbell, Taeusch, & Avery, 1976; and B. Silverman et al., 1998)

Placental transfer of glucose from the mother to the fetus occurs via facilitated diffusion across the placenta and therefore high circulating levels of blood glucose in the mother can significantly increase the amount of glucose that crosses the placenta. When the fetus is exposed to a hyperglycaemic in-utero environment, a response occurs. This involves increased secretion of fetal insulin to cope with the additional glucose load. This response can result in hyperinsulinaemia in the fetus. The added pressure placed on the fetus' pancreas in conjunction with high circulating glucose levels can damage the pancreatic beta cells and lead to long-term problems with carbohydrate metabolism after birth (Fetita, Sobngwi, Serradas, Calvo, & Gautier, 2006). The fetal response to maternal hyperglycaemia results in dysregulated metabolic programming, and a greater chance of the offspring developing metabolic complications in adulthood such as obesity, cardiovascular disease, and diabetes (B. Silverman, Metzger, Cho, & Loeb, 1995). In addition to the metabolic programming that

occurs as a result of the adaptation to the hyperglycaemic environment, the fetus has an increased likelihood of increased fat deposition due to excess fuel availability, which results in an increased risk of macrosomia (Ehrenberg, Mercer, & Catalano, 2004). If the baby is macrosomic (large for gestational age), it is particularly vulnerable to birth complications, including shoulder dystocia. Offspring of women who developed GDM are also more likely to experience cognitive and developmental impairment in infancy (Camprubi Robles et al., 2015).

The consequences of GDM are far-reaching, not just affecting the mother and her offspring but also placing pressure on the public healthcare system. The worldwide costs of non-communicable diseases including diabetes, metabolic syndrome, and CVD (consequences of GDM) have been estimated at upwards of 1.5 trillion dollars annually (Bloom et al., 2011). One approach to reducing the rates of non-communicable diseases is to target prevention strategies as early as possible, which could mean targeting women during pregnancy. Of particular concern is the contribution of GDM to the "vicious diabetes cycle" whereby the development of GDM is associated with an increased risk of both mother and offspring developing type 2 diabetes postnatally, and the offspring becoming obese in later life. Maternal obesity is a well-known risk factor for developing GDM during gestation, and therefore the cycle of diabetes repeats. Diabetes is one of the fastest growing non-communicable diseases in the modern world and every opportunity to reduce its growth, particularly its trans-generational expansion, needs to be explored.

The costs of GDM and its consequences are not just economical. As a result of the effects it has on health, the emotional and psychological impacts from the reduced quality of life that occurs with some of the long-term consequences, such as diabetes and cardiovascular diseases, are significant. An often-overlooked consequence of GDM is the effect of obesity on a child's social status and, in turn, their mental and psychological welfare (Puhl & Latner, 2007).

1.2.4 Identification of research need.

It is apparent that GDM is a pregnancy disorder that is growing in prevalence, in parallel with increasing rates of overweight and obese women becoming pregnant. An increasing prevalence of GDM results in an increased number of mothers and infants exposed to the detrimental consequences of GDM development. It should therefore be a public health priority to find solutions to a condition that increases disability, public healthcare costs,

morbidity, reduced quality of life, and societal burden (Y. Chen et al., 2009; Hex, Bartlett, Wright, Taylor, & Varley, 2012; Kolu, Raitanen, Rissanen, & Luoto, 2012; Langer et al., 2005; Lawson & Rajaram, 1994). The best way to minimise the associated consequences of GDM is to prevent its occurrence.

There are two main barriers to GDM prevention. Firstly, in order to deliver interventions to prevent GDM development, women at risk of developing GDM first need to be identified accurately. There is currently no accurate tool for predicting which women will go on to develop GDM, and the clinical risk-factors that are currently being used (age, BMI, ethnicity, previous GDM, family history of diabetes) fail to detect a large proportion of women who are eventually diagnosed with GDM. It is therefore important to investigate clinically feasible biomarkers for the detection of women who are likely to develop GDM. Secondly, once women have been identified as at-risk, it is necessary to investigate what strategies could be employed to change their disease risk trajectory and prevent GDM and its associated consequences. A promising strategy for the prevention of GDM, that is deemed safe, accessible, and well accepted by pregnant women, is modification of the maternal diet. An abundance of research has been conducted that links the maternal diet to GDM development and progression and this association should be explored further to understand how dietary modifications could be a means for preventative intervention strategies.

1.3 Maternal Diet and Gestational Diabetes Mellitus

1.3.1 Importance of diet during pregnancy.

Nutrition is vital for the maintenance of life. Nutrients which are ingested through food, beverages, and dietary supplements, provide the key ingredients to nurture life, and to sustain growth and development. The diet during pregnancy is particularly important as it is a time of metabolic change, hormonal change, and rapid growth. Pregnancy is a time where the mother not only needs to be consuming adequate nutrients to support her day-to-day functioning, but also needs to ensure adequate consumption to support a healthy pregnancy, optimal growth and development of her fetus, and preparation for lactation (Brown, 2011). The fetus is dependent on the mother's nutrient supplies in order to survive, grow, and flourish.

During pregnancy, the mother's blood volume and fat stores increase, new tissue grows (e.g., placenta), and the nutrient requirements of the fetus must be met. Table 1.2 lists some of the nutrients required in significantly higher amounts during pregnancy, an explanation for their

increased requirements, their recommended intake based on New Zealand food and nutrition guidelines for healthy pregnant and breastfeeding women, and the food-based guidelines provided for pregnant women to choose diet-based solutions to assist with meeting these requirements (Ministry of Health, 2008, 2014).

Table 1.2

Nutrient Recommendations During Pregnancy

Nutrient	Requirements in pregnancy	New Zealand Recommendations	New Zealand Food-based guidelines to meet requirements
Iron	For increased maternal blood volume during pregnancy. Extra requirements are offset by reduced blood loss due to menstruation in first trimester, but needs are increased in second and third trimesters (Bothwell, 2000).	Aged 14-50 years: Recommended daily intake = 27mg/day	Two serves per day of lean meats, chicken, seafood, eggs, cooked dried beans, peas and lentils, nuts or seeds.
Iodine	Required component of thyroid hormones. Increased needs in pregnancy due to rapid rates of growth in fetus and important for normal brain development (Delange, 2007).	Aged 14-50 years: Recommended daily intake = 220 µg/day	Well-cooked seafoods, milk, eggs, some cereals, seameal custard and commercially made bread. If using salt, choose iodised salt.
Folic Acid	Important for reducing the risk of neural tube defects (MRC Vitamin Study Research Group 1991).	800 µg (0.8mg) folic acid tablet daily for at least four weeks before and 12 weeks after conception. Women at increased risk of a pregnancy with a neural tube defect = 5000 µg (5 mg) tablet of folic acid daily for at least four weeks before and 12 weeks after conception.	Eat at least six servings of vegetables and fruit per day, aiming for 10 servings per day – at least four servings of vegetables and two servings of fruit.

Nutrient	Requirements in pregnancy	New Zealand Recommendations	New Zealand Food-based guidelines to meet requirements
Vitamin A	Essential for visual development and immune function. Excess Vitamin A intake, however, can be teratogenic (Azaïs-Braesco & Pascal, 2000).	Aged 19–50 years: Recommended daily intake = 800 µg of retinol equivalents Upper limit = 3000 µg retinol	Eat at least six servings per day of vegetables and fruit – at least four servings of vegetables and two servings of fruit.

1.3.2 Associations between maternal diet and gestational diabetes.

Diet has been a focus for interventions to prevent type 2 diabetes, manage symptoms, and reduce the adverse consequences associated with diabetes development (American Diabetes Association et al., 2008). With the strong association between diet and type 2 diabetes and the success of nutrition interventions in type 2 diabetes, the maternal diet has been manipulated for the management of gestational diabetes after diagnosis, and has been investigated for its role in the development of GDM. The maternal diet is a safe, inexpensive, modifiable lifestyle factor. Unlike medical-based interventions, it is more difficult to "overdose" or prescribe a highly detrimental diet, resulting in much less risk. Other than fat-soluble vitamins and excess energy intake, the human body has complex mechanisms that ensure that the over-consumption of most vitamins and minerals are excreted or removed safely from the body. The maternal diet as an intervention is also well received by pregnant women (Poston et al., 2013), whereas some women are apprehensive toward medical intervention over concern of potential risks to the unborn child, particularly in the early stages of drug development or early implementation (Widnes & Schjøtt, 2017).

An array of research has been conducted to investigate the role of the maternal diet in GDM development, with varied results. Associations have been found between individual micro- and macro-nutrients and GDM, combinations of nutrients, glycaemic index, food groups, and complete dietary patterns (Table 1.3). Due to the extensive nature of the research, the most recent research conducted (over the past 10 years) into the relationship between diet and the development of GDM is detailed in Appendix 1. Table 1.3 is a summary table of the nutrient/food relationships observed, with emphasis on the study design, sample size, and the direction of the observed relationship.

Table 1.3
A Summary Table of the Studies of Maternal Diet Related to GDM Between the Years 2007-2017

Dietary component	Direction of relationship in GDM compared to controls	Study design	Assessment	Number of Participants	Reference
Vitamin A	No difference	Observational	Blood levels	119	(Grissaa et al., 2007)
	Higher	Observational	Dietary intake (FFQ)	3787	(Salmenhaara et al., 2010)
Retinol	Lower	Observational	Blood levels	82	(Hekmat et al., 2014)
Vitamin B1	Lower	Observational	Blood levels	177	(Bartáková et al., 2016)
Vitamin B2	Lower	Observational	Dietary intake (3-day diary)	263	(S. Park et al., 2013)
Vitamin B12	No difference	Observational	Blood levels	785	(Krishnaveni et al., 2009)
	Lower	Observational	Blood levels	344	(Sukumar et al., 2016)
Vitamin C	Higher	Observational	Blood levels	119	(Grissaa et al., 2007)
		Observational	Dietary intake (3-day diary)	263	(S. Park et al., 2013)
Vitamin D	Lower	RCT	-	92	(Shahgheibi et al., 2016)
		Observational	Blood levels	402	(Zuhur et al., 2013)
		Observational	Blood levels	400	(O. Wang et al., 2012)
		Observational	Blood levels	335	(Parlea et al., 2012)
		Observational	Blood levels	1953	(Zhou et al., 2014)
		Observational	Blood levels	90	(Haidari et al., 2016)
		Observational	Blood levels	2320	(Dodds et al., 2016)
		Observational	Blood levels	709	(Lacroix et al., 2014)
		RCT	-	470	(Mojibian et al., 2015)

Dietary component	Direction of relationship in GDM compared to controls	Study design	Assessment	Number of Participants	Reference
		Observational	Blood levels	4090	(Schneuer et al., 2014)
		Observational	Blood levels	399	(McLeod et al., 2012)
		Observational	Blood levels	180	(Baker et al., 2012)
		Observational	Blood levels	1100	(Savvidou et al., 2011)
		Observational	Blood levels	248	(Makgoba et al., 2011)
		Observational	Blood levels	171	(C. Zhang et al., 2008)
		Observational	Blood levels	593	(Farrant et al., 2009)
		Observational	Blood levels	2358	(Rodriguez et al., 2015)
		Observational	Blood levels	524	(Kramer et al., 2014)
		Observational	Blood levels	136	(Domaracki et al., 2016)
	No difference	Observational	Blood levels	1544	(Boyle et al., 2016)
		RCT	-	179	(Yap et al., 2014)
		Observational	Blood levels	940	(Loy et al., 2015)
		Observational	Blood levels	708	(Arnold et al., 2015)
		Observational	Blood levels	76	(Pleskačová et al., 2015)
		Observational	Blood levels	523	(S. Park et al., 2014)
		Observational	Blood levels	240	(Flood-Nichols et al., 2015)
		Observational	Blood levels	1467	(Whitelaw et al., 2014)
		Observational	Blood levels	149	(Nobles et al., 2015)
	Higher	Observational	Dietary intake (FFQ)	3787	(Salmenhaara et al., 2010)

Dietary component	Direction of relationship in GDM compared to controls	Study design	Assessment	Number of Participants	Reference
Vitamin E	Lower	Observational	Blood levels	119	(Grissaa et al., 2007)
		Observational	Dietary intake (FFQ)	195	(Ley et al., 2013)
Iron (non-haem)	Lower	Observational	Dietary intake (FFQ)	7229	(Darling et al., 2016)
Iron (haem)	Higher	Observational	Dietary intake (FFQ)	13475	(Bowers et al., 2011)
		Observational	Dietary intake (FFQ)	3158	(Qiu et al., 2011)
Iron	No difference	Observational	Ferritin levels	699	(Bowers et al., 2016)
		RCT	-	1042	(Chan et al., 2009)
		Observational	Ferritin levels	104	(Zein et al., 2015)
	Higher	Observational	Ferritin levels	3776	(Khambalia et al., 2015)
		Observational	Ferritin & Hepcidin levels	321	(Rawal et al., 2017)
		Observational	Ferritin levels	200	(Omidvar et al., 2013)
		Observational	Blood levels	1033	(Behboudi-Gandevani et al., 2013)
		Observational	Ferritin & hepcidin levels	102	(Derbent et al., 2013)
		Observational	Dietary intake (FFQ)	399	(Helin et al., 2012)
		Observational	Dietary intake (FFQ)	3787	(Salmenhaara et al., 2010)
	Lower	Observational	Blood levels	60	(Akhlaghi et al., 2012)
Calcium	Higher	Observational	Blood levels	1467	(Whitelaw et al., 2014)
	No difference	Observational	Dietary intake (FFQ)	3414	(Osorio-Yáñez et al., 2016)
	Lower	Observational	Dietary intake (3-day diary)	263	(S. Park et al., 2013)
Sodium	Lower	Observational	Dietary intake (3-day diary)	263	(S. Park et al., 2013)

Dietary component	Direction of relationship in GDM compared to controls	Study design	Assessment	Number of Participants	Reference
Cholesterol	Lower	Observational	Blood levels	119	(Grissaa et al., 2007)
	Higher	Observational	Dietary intake (FFQ)	13475	(Bowers et al., 2012)
		Observational	Dietary intake (FFQ)	3898	(Qiu et al., 2011)
Folate	No difference	Observational	Blood levels	785	(Krishnaveni et al., 2009)
		Observational	Blood levels	344	(Sukumar et al., 2016)
	Higher	Observational	Supplement Questionnaire	1938	(Zhu et al., 2016)
		Observational	Dietary intake (FFQ)	3787	(Salmenhaara et al., 2010)
Zinc	No difference	Observational	Blood levels	1033	(Behboudi-Gandevani et al., 2013)
		Observational	Blood levels	62	(Hamdan et al., 2014)
Selenium	No difference	Observational	Blood levels	62	(Hamdan et al., 2014)
	Lower	Observational	Blood levels	131	(Kilinc et al., 2008)
Magnesium	Lower	Observational	Blood levels	85	(Goker Tasdemir et al., 2015)
Myo-inositol	Lower	RCT	-	220	(D'Anna et al., 2015)
		Clinical trial	-	83	(D'Anna et al., 2012)
	Higher	Observational	Urine levels	94	(Murphy et al., 2016)
		Observational	Amniotic fluid	60	(Santamaria et al., 2016)
Salty food cravings	Lower	Observational	Questionnaire	2022	(Farland et al., 2015)
Animal fat	Higher	Observational	Dietary intake (FFQ)	13475	(Bowers et al., 2012)
Mono-unsaturated fatty acids	Higher	Observational	Dietary intake (FFQ)	13475	(Bowers et al., 2012)
		Observational	Blood levels	147	(Chen et al., 2010)

Dietary component	Direction of relationship in GDM compared to controls	Study design	Assessment	Number of Participants	Reference
Saturated fat	Higher	Observational	Dietary intake (24-hr recall)	746	(H. -J. Park et al., 2013)
		Observational	Dietary intake (3-day diary)	263	(S. Park et al., 2013)
	Lower	Observational	Dietary intake (FFQ)	3787	(Salmenhaara et al., 2010)
Energy intake	Higher	Observational	Dietary intake (24-hr recall)	746	(H. -J. Park et al., 2013)
		Observational	Dietary intake (3-day diary)	263	(S. Park et al., 2013)
Fat as a percentage of energy intake	Lower	Observational	Dietary intake (24-hr recall)	746	(H. -J. Park et al., 2013)
		Observational	Dietary intake (FFQ)	205	(Ley et al., 2011)
Carbohydrate as a percentage of energy intake	Lower	Observational	Dietary intake (FFQ)	205	(Ley et al., 2011)
Fibre	Higher	Observational	Dietary intake (FFQ)	3787	(Salmenhaara et al., 2010)
Protein	Higher	Observational	Dietary intake (FFQ)	3787	(Salmenhaara et al., 2010)
Animal protein	Higher	Observational	Dietary intake (FFQ)	21457	(Bao et al., 2013)
Vegetable protein	Lower	Observational	Dietary intake (FFQ)	21457	(Bao et al., 2013)
Sugar	Lower	Observational	Dietary intake (FFQ)	3787	(Salmenhaara et al., 2010)
Eggs	Higher	Observational	Dietary intake (FFQ)	3898	(Qiu et al., 2011)
Potatoes	Higher	Observational	Questionnaire	21693	(Bao et al., 2016)
Fried food	Higher	Observational	Dietary intake (FFQ)	21079	(Bao et al., 2014)
Red meat	Higher	Observational	Dietary intake (FFQ)	21457	(Bao et al., 2013)
Nuts	Lower	Observational	Dietary intake (FFQ)	21457	(Bao et al., 2013)
Leafy green vegetables, fruit,	Lower	RCT	-	1008	(Sahariah et al., 2016)

Dietary component	Direction of relationship in GDM compared to controls	Study design	Assessment	Number of Participants	Reference
and milk snack					
Milk products, cereal products, vegetables, meat	Higher	Observational	Dietary intake (FFQ)	3787	(Salmenhaara et al., 2010)
High-protein foods and sugar-sweetened beverages, dairy products, smoked meat, pork, poultry, fresh fruit, syrup, and goodies	Higher	Observational	Dietary intake (FFQ)	363	(Bartáková et al., 2016)
Fresh vegetables	Lower	Observational	Dietary intake (FFQ)	363	(Bartáková et al., 2016)
Mediterranean diet adherence	Lower	Observational	Dietary intake (FFQ)	1003	(Karamanos et al., 2014)
	No difference	Observational	Dietary intake (24-hr recall)	463	(Izadi et al., 2016)
DASH diet	Lower	Observational	Dietary intake (24-hr recall)	463	(Izadi et al., 2016)
Low-carbohydrate diet	Higher	Observational	Dietary intake (FFQ)	21411	(Bao et al., 2014)
Glycaemic index	No effect	RCT	-	125	(Markovic et al., 2016)
High-GI fruits and energy-dense snacks	Higher	Observational	Dietary intake (FFQ)	571	(Li et al., 2015)
Coffee	Lower	Observational	Dietary intake (3-day diary)	263	(S. Park et al., 2013)
Night-time eating	Higher	Observational	Dietary intake (3-day diary)	263	(S. Park et al., 2013)
	No difference	Observational	Questionnaire	1061	(Loy et al., 2017)
Diet quality	No difference	Observational	Dietary intake (FFQ)	137	(Meinila et al., 2017)
		Observational	Dietary intake (FFQ)	1907	(Gresham et al., 2016)
Omega-3	Higher	Observational	Dietary intake (FFQ)	1733	(Radesky et al., 2007)

Despite the considerable amount of evidence for an association between maternal diet and GDM, the findings are not cohesive and as such the details of this association remain inconclusive. There are some possible reasons for the inconsistent findings. Firstly, the majority of studies conducted have relied on observational findings. Observational findings are often confounded by many other aspects of free-living participants' situations, including age, BMI, ethnicity, and socioeconomic status, that need to be taken into account during analysis. Observational studies also rely heavily on self-reported dietary assessment. Self-reported dietary assessment includes food frequency questionnaires (FFQs), 24-hr recalls, and food diaries, all of which are subject to multiple biases and can only ever provide a subjective insight into dietary exposure. One example of bias in dietary assessment methods is social desirability bias, where the participant reports eating a diet that is perceived as favourable/healthy, driven by the subconscious intention of appearing healthy to the researcher. Another common limitation of self-reported dietary assessment is recall bias, particularly in 24-hr recalls or FFQs which are collected retrospectively. Retrospective recall of the details of dietary consumption can be easily forgotten or estimated inaccurately. Common dietary assessment methods also do not provide information as to *how* and *why* a diet is related to disease. Observational studies rely on drawing hypotheses to explain the associations from subjective dietary assessments, without objectively investigating their role in metabolism.

A further consideration is the effect of the gut microbiome, which can play an important role in the association between the diet and disease interaction. Unfortunately, many studies find it difficult to recruit participants who are willing to donate faecal samples, or are limited in their budgets to collect and analyse such samples, leaving this important aspect of the association between disease and diet relatively under-explored. With recognition of the importance of the gut flora in health and disease, studies in this area are increasing markedly, and may provide important insight into understanding the association between maternal diet and GDM. One study has found that the administration of a probiotic supplement, containing *Lactobacillus rhamnosus* GG and *Bifidobacterium lactis* BB-12 strains, in early pregnancy, significantly reduced the rates of women who developed GDM, furthering the evidence for the important role of the gut microbes in the development and progression of GDM (Luoto, Laitinen, Nermes, & Isolauri, 2010). Further investigations are warranted, but fall outside the scope of this thesis.

In addition to the limitations of the typical dietary assessment techniques used in dietary association studies, there is also a fundamental oversight in that nutrients are often analysed independently and treated as individual entities, rather than as small parts of a whole. Nutrients are not consumed in isolation in a typical human diet and arguably, should not be analysed as such. Nutrient combinations, as well as the food matrix they are consumed in, have an important effect on how nutrients are retained, absorbed, and utilised by the body (Jacobs & Tapsell, 2007; Parada & Aguilera, 2007). It is accepted that in some cases a food item or meal is much more important than the sum of its parts. An example of the effect of nutrient combinations is glycaemic index. When foods are consumed in combination, the glycaemic load of a meal can be dramatically different to the glycaemic index of the individual foods. The individual food could have a high glycaemic index, but when consumed in combination with certain foods, the overall meal will have a reduced glycaemic load. One way to account for the effect of consuming nutrients in combination is to analyse patterns of dietary intake rather than single nutrients or foods.

1.3.3 Dietary pattern analysis.

Dietary pattern analysis is an approach to dietary investigations that involves the analysis of the sum of foods and beverages consumed over a given period, and therefore takes into consideration nutrients consumed in combination (Hu, 2002). This type of analysis takes into account that combinations of nutrients can have a greater impact on health and disease than the sum of their parts. Analysing dietary patterns also provides a more practical approach to translating research findings into guidelines, as it is closer to the typical dietary behaviour of free-living participants than isolated nutrient analysis.

There are three statistical approaches which are most commonly used for determining dietary patterns in an a posteriori way, where the dietary data is used to generate dietary patterns without preconceived ideas or influences (Hoffmann, Schulze, Schienkiewitz, Nöthlings, & Boeing, 2004; Reedy et al., 2009); factor analysis, cluster analysis, and reduced rank regression. Factor analysis and cluster analysis are the most commonly used methods to derive dietary patterns, whereas reduced rank regression is typically used when dietary assessment can be combined with biological response markers.

Factor analysis (often used synonymously with principal component analysis in dietary pattern research) is used to produce factors containing correlated observed variables (the dietary variables) which represent latent variables (Reedy et al., 2009). In the case of dietary

pattern analysis, the latent variables represent different food consumption patterns. Each food group is given a factor loading for each factor, which is the correlation between that food group and the factor. A score on each factor is calculated for each participant by summation of the products (multiplication results) of the factor loadings with the participant's intake of that food group.

Cluster analysis is another statistical approach to deriving dietary patterns where individuals are clustered together based on the frequency of food consumed in the groups, food quantity, and the contribution of the different food groups to overall energy consumption (Reedy et al., 2009). The main difference between factor analysis and cluster analysis is that factor analysis groups food groups together, whereas cluster analysis groups individuals together. Cluster analysis results in independent and mutually exclusive clusters, whereas factor analysis factors can produce groups that are highly correlated. Orthogonal rotation in factor analysis is required to ensure that factors are mutually exclusive and multicollinearity is removed, allowing for appropriate interpretation of downstream analyses, which are often regression-based (Newby, Muller, Hallfrisch, Andres, & Tucker, 2004). Cluster analysis is, however, very sensitive to outliers (Crozier, Robinson, Borland, Inskip, & SWS Study Group, 2006; Reedy et al., 2009).

Reduced rank regression is used when combining dietary data with response variables that are related to the outcome of interest (Hoffmann, Schulze, Schienkiewitz, Nöthlings, & Boeing, 2004). In order to use reduced rank regression, robust dietary-related biological responses associated with the outcome of interest need to first be identified from existing literature. The use of a priori information makes reduced rank regression an only partially empirical approach to derive dietary patterns.

Studies investigating the association between maternal diet and GDM, using statistically derived dietary pattern analysis, are listed in Table 1.4 along with their key findings.

Table 1.4

Studies of Maternal Dietary Patterns and GDM Development

Authors & Date of Publication	Participants	Dietary Assessment Method	Major Findings
C. Zhang et al. (2006)	13,110 participants who reported at least one singleton pregnancy in the Nurses' Health Study II, United States of America 758 diagnosed with GDM	Validated pre-pregnancy semi-quantitative food frequency questionnaire (FFQ)	Two dietary patterns ("prudent" and "Western") were identified in the pre-pregnancy cohort using factor analysis. Prudent pattern = high intake of fruit, green leafy vegetables, poultry and fish. Western pattern = high intake of red meat, processed meat, refined grain products, sweets, French fries and pizza. The Western dietary pattern was significantly associated with an increased risk of developing GDM. Red meat and processed meat consumption pre-pregnancy individually contributed to an increased risk of GDM.
Radesky et al. (2007)	1733 women with singleton pregnancies enrolled in Project Viva, United States of America 91 diagnosed with GDM 206 women had impaired glucose tolerance (IGT)	First trimester diet assessed using a validated FFQ	Two dietary patterns were identified in the cohort using principal components factor analysis: A prudent dietary pattern (high in vegetables, fruit, legumes, fish, poultry, eggs, salad dressing and whole grains) and a Western dietary pattern (high in red and processed meats, sugar-sweetened beverages, French fries, high-fat dairy products, desserts, butter and refined grains). No dietary patterns in the study were associated with GDM outcome.

Authors & Date of Publication	Participants	Dietary Assessment Method	Major Findings
He et al. (2015)	3063 pregnant Chinese women 544 diagnosed with GDM	Validated FFQ conducted at time of diagnosis (24-27th gestational week)	Four dietary patterns were identified using factor analysis. The "vegetable" dietary pattern (high in root vegetables, beans, mushrooms, melon vegetables, seaweed, other legumes, fruits, leafy and cruciferous vegetables, processed vegetables, nuts, and cooking oil) was associated with a reduced likelihood of GDM. The "sweets and seafood" dietary pattern, which was high in Cantonese desserts, molluscs and shellfish, and sugar-sweetened beverages, and low in leafy and cruciferous vegetables and grains, which were mainly refined, was associated with an increased likelihood of GDM.
Tryggvadottir et al. (2016)	168 women from Iceland 17 diagnosed with GDM	4-day weighed food record (commencement between 19th and 24th weeks of gestation)	Two dietary patterns were identified using principal component analysis; only one was associated with GDM. The dietary pattern associated with reduced risk of GDM consisted of positive loadings for seafood, eggs, vegetables, fruit and berries, vegetable oils, nuts and seeds, pasta, breakfast cereals, and coffee and tea, and negative loadings for soft drinks and French fries.
Schoenaker et al. (2015)	3853 women from Australia 292 diagnosed with GDM	Validated 101-item FFQ conducted pre-pregnancy	Four dietary patterns identified using factor analysis: A "meats, snacks and sweets" pattern characterised by high consumption of red and processed meat, cakes, sweet biscuits, fruit juice, chocolate and

Authors & Date of Publication	Participants	Dietary Assessment Method	Major Findings
			pizza. A “Mediterranean-style” pattern characterised by high consumption of vegetables, legumes, nuts, tofu, rice, pasta, rye bread, red wine and fish. A “fruit and low-fat dairy” pattern characterised by high consumption of fruits and low-fat dairy including yoghurt, low-fat cheese and skimmed milk. A “cooked vegetables” pattern characterised by high consumption of carrots, peas, cooked potatoes, cauliflower and pumpkin. Higher scores on the Mediterranean-style dietary pattern were significantly associated with a reduced likelihood of developing GDM, after adjustment for confounding variables.
Shin et al. (2015)	249 women from the NHANES study, United States of America 34 diagnosed with GDM	24-hr dietary recall (50% first trimester, 27% second trimester, 23% third trimester)	Three dietary patterns were identified using reduced rank regression: A “high refined grains, fats, oils and fruit juice” pattern; a “high nuts, seeds, fat and soybean; low milk and cheese” pattern; and a “high added sugar and organ meats; low fruits, vegetables and seafood” pattern. All three dietary patterns were associated with an increased risk of GDM, after adjustment for confounding variables.

Each of the dietary pattern studies found slightly different results, as would be expected when comparing across different populations of free-living participants. However, when considered as a whole, the general findings were that a diet that contained higher amounts of fresh fruit, vegetables, lean meats, and wholegrains was associated with a reduced risk of GDM, and a dietary pattern higher in processed and red meats, sugar-sweetened beverages, deep-fried potatoes, and burgers was associated with an increased risk of GDM. However, the majority of dietary pattern studies performed used participants in Western populations. Only one study to date has conducted dietary pattern investigations related to GDM in an Asian population (He et al., 2015). With the rates of GDM in South-East Asia being some of the highest worldwide (estimated at 24% of all live births having been exposed to maternal hyperglycaemia during pregnancy; IDF, 2015), it is clear that this is an at-risk population that requires intervention. Therefore, understanding the role that the local diet plays in GDM development in this population is very important. The dietary patterns generated by He et al. (2015) were distinct from those reported in studies conducted in Western populations, demonstrating the uniqueness of typical dietary consumption in Asia. This uniqueness implies a need for further studies in Asian populations, and demands a cautious approach to the translation of dietary findings across continents and cultures. Although dietary pattern analysis addresses the concerns around analysing nutrients as individual entities, it does not address the limitations of self-reported, subjective dietary assessment – dietary patterns are derived from information obtained from these assessments.

1.3.4 Identification of research need.

Despite the abundance of research conducted to investigate the role of the maternal diet in GDM development, there remains a lack of agreement as to what components of the diet are contributing to increased risks of GDM. There is also a lack of studies that have experimentally underpinned the explanation for *how* the maternal diet is related to GDM development. The majority of studies have been conducted using an individual nutrient approach, whereby observational findings of dietary associations with GDM development have been attributed to an individual nutrient in the diet. A dietary pattern approach is the way forward for investigating associations between diet and disease in free-living participants, where the findings can be directly transposed to food-related recommendations. The dietary pattern associations which have been generated in relation to GDM so far have predominantly been conducted in Western populations and there is a great need for understanding dietary associations with GDM in an Asian context, where GDM prevalence is

some of the highest worldwide. Additionally, few studies have been conducted investigating both dietary data and biological data to link the subjective observational dietary associations to metabolic mechanisms in order to discover *how* and *why* the diet is associated with GDM. Investigating observational dietary associations from a metabolic perspective would assist with understanding the role of diet in disease development and progression, which would better inform targeted nutritional interventions aimed at preventing the metabolic dysfunction that leads to GDM.

1.4 Metabolomics and Gestational Diabetes Mellitus

1.4.1 Definition of metabolomics and its applications.

Metabolomics is the study of low molecular weight compounds (Oliver, Winson, Kell, & Baganz, 1998). The metabolome is the comprehensive set of compounds found in a system of interest and includes both endogenous and exogenous compounds. Endogenous compounds include amino acids, fatty acids, nucleic acids, hormones, enzymes, amines, vitamins, organic acids, short peptides, and sugars, whereas exogenous compounds include those derived from environmental contaminants, food additives, toxins, and drug metabolites (Wishart et al., 2013). The metabolome is downstream from genomics, transcriptomics, and proteomics and is influenced by diet, lifestyle, environment, and physical activity, thus making the metabolome the closest representation of the phenotype of an individual. Over 40,000 metabolites have been reported for the human metabolome (Wishart et al., 2013).

There are two approaches to investigating the metabolome, an untargeted and a targeted, quantitative approach (Patti, Yanes, & Siuzdak, 2012). By definition, metabolomics is an untargeted approach, where the metabolome is characterised without a priori information, aiming to detect and identify as many compounds as possible with the experimental method employed (Nicholson & Lindon, 2008). Bias in metabolomics can still occur because no single method can identify the full metabolome. A combination of analytical platforms and sample preparations are required. A targeted “metabolomics” approach, in contrast, deals with fewer compounds and the experimental methods are designed specifically to detect the compounds of interest. A targeted approach may be appropriate in a context where clear understandings of the metabolic pathways of interest are already well established. This is a bottom-up approach to investigating the metabolome, whereas an untargeted, data-driven, top-down approach allows for hypothesis generation and more discovery-based research, which is particularly important in areas where little is known or certain. Metabolomics has

been used as an analytical approach in the research areas of pharmacy, food science, medical science, microbiology, and nutrition, where its applications have included biomarker discovery, drug detection, food contamination, mechanistic studies, drug safety, food toxicity, differentiation of disease state from normal, and the assessment of dietary interventions (Inoue et al., 2016; Li et al., 2015; Pamungkas, Park, Lee, Jee, & Park, 2016; S. Park et al., 2013; Stradling et al., 2016).

1.4.2 Methodologies in metabolomics.

The two main analytical methods that have been employed to investigate the metabolome are mass spectrometry-based (MS) and nuclear magnetic resonance spectroscopy (NMR) methods (Pan & Raftery, 2007). Both NMR and MS have advantages and limitations (Appendix 2), and the choice of instrument depends on the research question, access, time, budget, and expertise available. No single technique can detect the complete range of low-weight compounds in a system, and as such a combination of approaches is required to analyse the metabolome comprehensively. Mass spectrometry-based metabolomics is often combined with separation techniques such as liquid chromatography (LC), gas chromatography (GC), or capillary electrophoresis (CE) to separate the compounds prior to identification, which requires further chemical modification in order to extract metabolites from biological samples and prepare them to be run through the GC or LC columns, or CE capillaries prior to MS (Raterink et al., 2014).

1.4.3 Specimens for investigation in human metabolomics.

For human-based metabolomics studies, there are a range of biological specimens that can be analysed. The choice of specimen depends on the research question and the intended outcomes from the findings (Horgan et al., 2009). For example, it would be inappropriate to analyse tissue samples from a live pregnant woman's organs, a highly invasive and potentially risky procedure, if the intended outcome was to develop a clinical screening tool. However, placental samples or umbilical cord blood obtained after delivery can be a valuable source of information related to maternal-fetal status and for the purpose of studying biological mechanisms of pregnancy complications (Fanos et al., 2013). The range of human specimens suitable for analysis include plasma, serum, hair, urine, amniotic fluid, cord blood, faeces, skin swabs, breath, saliva, and tissues. Each sample contains different information about the condition or effect of an intervention. Table 1.5 provides a list of the advantages and limitations of the most commonly analysed types of human specimen. Urine and blood

samples are the most commonly analysed biological specimens in medical metabolomics as they are often collected routinely for other clinical tests and evaluations. Urine is also non-invasive to collect, and can be collected in large volumes. However, urine represents breakdown and waste products of a system and can be easily influenced by recent dietary intake or lifestyle changes and is better suited to assessing short-term exposures or real-time metabolic state (Gibney et al., 2005).

Human blood is preferable to urine for detecting longer term changes in metabolism and, in developed countries, blood-sample collection is also relatively safe, and minimally invasive. Human blood is a suitable biofluid to collect as it represents the overall metabolism of an individual, being influenced by multiple organs, as well as the metabolic effects of external exposures (Gowda et al., 2008). It is also highly regulated and any shifts in its metabolic profile are therefore very meaningful, and provide important information. Blood is less affected by external influences, such as non-nutrient metabolites resulting from cooking processes and recent dietary intake, than urine, making it less subject to confounding external variables (Gibney et al., 2005). Human blood is typically processed and stored as serum and plasma. Processing can involve the addition of an anticoagulant, and storage in plastic containers. Addition of anticoagulant and exposure to plasticizers can alter the chromatographic profile of the sample, and should be taken into account when analysing such samples (Gonzalez-Covarrubias et al., 2013; Yin, Lehmann, & Xu, 2015).

Amniotic fluid and maternal tissue (e.g., liver, placenta, or pancreatic tissue) are exceptionally invasive to collect, and collection poses risks to both mother and fetus (Mujezinovic & Alfirevic, 2007). As such, human tissue specimens are less desirable for identifying clinically useful biomarkers, but maintain an important role in investigations related to the elucidation of metabolic mechanisms of disease.

Recently, interest in the role of the gut microbiota on the development and progression of disease has led to an increase in metabolomic analysis of faecal matter. Metabolomic analysis of faeces can provide information on the composition of the gut microbiome and how they are responding to particular diets or interventions. However, it is much more difficult to recruit participants willing to provide a faecal sample, in comparison to urine or a blood sample (Lecky, Hawking, & McNulty, 2014).

Table 1.5

Advantages and Limitations of the Common Biological Specimen Types Analysed in Human Metabolomics Research

Specimen	Advantages	Disadvantages	Potential Uses
Blood (plasma & serum)	• Safe to collect • Overview of long-term metabolism • Blood samples are commonly collected in clinical practice = easier access to samples	• Requires phlebotomist to collect • Difficult to determine origin of findings • Exposure to blood-borne diseases	• Assess effect of exposures, interventions, and changes to metabolic homeostasis
Urine	• Non-invasive • Sterile • Easy to collect	• Influenced by effect of gut microbes and cooking methods and daily diet fluctuations	• Assess short-term effect of exposures and interventions
Faeces	• Ability to assess effects of gut microbes on health	• Odour • Can be difficult to recruit participants willing to collect samples	• Assess dietary exposure and actions of gut microbes
Tissues	• Useful for mechanistic studies • Can use to evaluate direct effects of an intervention on a tissue/organ of interest	• Invasive • Not suitable for clinical screening tools where healthy women are used	• Identify tissue-specific mechanisms of action
Amniotic fluid	• Useful for mechanistic studies • Ability to assess fetal metabolome while still in-utero	• Invasive • Requires highly trained medical professionals to collect • Not suitable for clinical screening tools where healthy women will be in the screening process	• Assess fetal metabolism and some maternal-fetal exchange

1.4.4 Metabolomics in gestational diabetes mellitus research.

Metabolomics is an established analytical method that has potential applications in GDM research. Metabolomics has proven useful as a tool for differentiating women at risk of pregnancy disorders from healthy controls, in the areas of fetal growth restriction and pre-eclampsia (Horgan et al., 2011; Kenny et al., 2010). As discussed in Section 1.2.2, there are currently no clinically accurate tools to predict women who are at risk of developing GDM, limiting options for prevention strategies. Typically, GDM intervention occurs after diagnosis in later pregnancy, when the detrimental consequences of maternal hyperglycaemia are already underway. Studies detailed in Table 1.6 employed metabolomic techniques to search for biomarkers of GDM development, with varying results. Overall, the studies conducted thus far have not managed to identify and validate any robust biomarkers of GDM.

Table 1.6
Studies Conducted to Assess the Effect of Gestational Diabetes on the Metabolome

Authors & year of publication	Specimen analysed & participant numbers	Analytical platform	Increased in GDM	Decreased in GDM
Graça et al. (2010)	Second trimester amniotic fluid Cases: N= 27 Controls: N= 82	NMR	Glucose	Glutamate, Glycine, Proline, Serine, Taurine, Acetate, Formate, Creatinine, Glycerophosphocholine
Diaz et al. (2011)	Second trimester Urine: Cases: N= 29 Controls: N= 25 Blood: Cases: N= 14 Controls: N= 20	NMR	Urine: 2-hydroxyisobutyrate, 3-hydroxyisovalerate, Choline, N-methyl-2-pyridone-5-carboxamide, N-methyl-nicotinamide	Blood: Betaine, Trimethylamine-N-oxide (TMAO)
Graça et al. (2012)	Second trimester Urine: Cases: N= 20 Controls: N= 21 Amniotic Fluid: Cases: N= 23 Controls: N= 26	Ultra-high performance liquid chromatography coupled to mass spectrometry (UPLC-MS)	Urine: Choline Amniotic fluid: A hexose (mild)	Amniotic Fluid: Glutamate (mild)
Sachse et al. (2012)	Urine at two time points of interest (8-20 weeks gestation and 28 weeks +/- 2 weeks) Cases: N=107 Controls: N= 716	NMR	Early: Glucose, Lysine, Citrate Later: Glucose, Citrate	

Authors & year of publication	Specimen analysed & participant numbers	Analytical platform	Increased in GDM	Decreased in GDM
Diaz et al. (2013)	Second Trimester Urine Cases: N=42 Controls: N= 84	NMR	Choline, Glucose, N-methyl-nicotinamide, Xylose	4-HPA, Hippurate
Dudzik et al. (2014)	Second Trimester Plasma and Urine Cases: N=20 Controls: N=20	LC-QTOF/MS, GC-Q/MS, and CE-TOF/MS	Plasma: 2-hydroxybutyric acid, Glycerol, 3-hydroxybutyric acid, Linoleic acid, Fumaric acid Urine: Histidine	Plasma: LPE (20:2), LPE (20:1), Trihydroxy-cholestanoyl taurine, LPA (18:2), LPC (20:5) sn-2, LPI(20:4), LPC(18:2) sn-2, PC(21:1), LPC(18:1) sn-2, LPE(22:4), LPS(20:0), Lipoxin C4, LPC(22:5) sn-2, LPC(22:4) sn-2, LPI(18:2), LPC(20:2) sn-2, LPC(20:4) sn-2, Taurolithocholic acid glucuronide, LPC(19:1), Glycerophosphocholine, DHA methyl ester, LPI(22:6), Arachidonic acid methyl ester, LPC(22:6) sn-2, LPI(20:3), LPE(18:2), LPE(20:4), LPE(22:6), Creatinine, Pyruvic acid, L-tryptophan, Glycine, L-glutamic acid, Lauric acid Urine: Carnitine
de Seymour et al. (2014)	Second trimester serum Cases: N= 22 Controls: N= 26	GC-MS	Itaconic Acid	

Authors & year of publication	Specimen analysed & participant numbers	Analytical platform	Increased in GDM	Decreased in GDM
Pinto et al. (2015)	Plasma Pre-diagnosis of GDM (16-24 weeks): Cases: N = 49 Controls: N = 32 Post-diagnosis of GDM (24-27 weeks): Cases: N = 12	NMR	Pre-diagnosis: Valine, Pyruvate	Pre-diagnosis: Proline, Urea, 1,5-anhydroglucitol Post-diagnosis: Glutamine, Creatinine, Betaine, Proline. Methanol, Glucose, Valine, Pyruvate, Alanine, Trimethylamine N-oxide (TMAO)
Liu et al. (2015)	Third trimester serum Cases: N= 12 Controls: N= 10	UPLC/Q-TOF-MS	1-methyladenosine, homovanillic acid, trans-3-octenedioic acid, glucosamine, LysoPC(16:1), 3-methylthiopropionic acid, LysoPC(20:2), Lactic acid, Tyrosine, 2-hydroxyglutaric acid, Phosphorylcholine	Gluconolactone, serinyl-glycine, LysoPC(20:4(5Z,8Z,11Z,14Z)), 3-methylhistidine, 2-oxo-4-methylthiobutanoic acid, SM (d18:0/18:0), indoleacetaldehyde, PE(14:1/20:0), 3-oxooctadecanoic acid, Glycine, Pyridoxal, Capric acid, Stearic acid, 3-methylhistidine, Arabonic acid
He et al. (2016)	Third trimester hair Cases: N= 47 Controls: N= 47	GC-MS	Adipic acid	

Authors & year of publication	Specimen analysed & participant numbers	Analytical platform	Increased in GDM	Decreased in GDM
Law et al. (2017a)	First, second, and third trimester urine Cases: N=27 Controls: N= 34	UPLC-MS	Oxitriptan (5-hydroxytryptamine), Serotonin, 5-hydroxykynurenamine, Indoleacetaldehyde, Indole-3-acetamide, Indoleacetic acid, 2-aminobenzoic acid, 4-(2-aminophenyl)-2,4dioxobutanoic acid, 2-aminomuconic acid, Semialdehyde, (R)-3-[(R)-3-hydroxybutanoyloxy]butanoate, 7-aminomethyl-7-carbaguanine, Imidazoleacetic acid, Hypoxanthine, Xanthine, Xanthosine, 1-methylhypoxanthine, 7-methylguanine, 1-methyladenosine, N4-acetylcytidine, Fusarin C Cis-1,2-dihydroxy-1,2-dihydro-7-methylnaphthalene (second and third trimester only) 3,4-methylenesebacic acid (first and second trimesters only) 16α-hydroxydehydroisoandrosterone (in second and third trimesters only) Cortolone-3-glucuronide (first trimester only)	

Authors & year of publication	Specimen analysed & participant numbers	Analytical platform	Increased in GDM	Decreased in GDM
Dudzik et al. (2017)	Second and third trimester plasma Cases: N= 24 Controls: N= 24	GC-Q/MS	Second trimester: Glycerol, 2-hydroxybutyrate, 3-hydroxybutyrate, Linoleic acid, Oleic acid, Palmitoleic acid, Palmitic acid Third trimester: Glycerol, Palmitic acid, Oleic acid	
Law et al. (2017)	First, second, and third trimester plasma Cases: N= 27 Controls: N= 34	UPLC-MS		A large range of unsaturated phospholipids (see original manuscript for details) However, if adjusted for multiple comparison testing, likely that no metabolites would remain statistically significant.

Two studies have investigated the amniotic fluid metabolome related to GDM (Graça et al., 2010, 2012). Graça et al. (2010) found statistically significant differences in nine metabolites between cases with pre-diagnostic GDM and controls with uncomplicated pregnancies, using NMR. The differences included decreased levels of glutamate, glycine, proline, serine, taurine, acetate, formate, creatinine, and glycerophosphocholine in cases, and increased glucose. These findings were not replicated in their later study using an untargeted metabolomic approach with UPLC-MS (Graça et al., 2012). Compared to other specimen choices, amniotic fluid is much more invasive and can pose risk to the pregnancy (Anandakumar et al., 1992). It should be discouraged unless the benefits are validated and heavily outweigh the risks. It is not feasible for amniotic fluid to be used in a universal screening setting.

Diaz et al. (2011) performed the first exploratory study on the urine metabolome in GDM, using samples collected in second trimester. Diaz et al. (2011) identified five metabolites that differed significantly between case and control groups. These included metabolites thought to play a role in biotin metabolism, amino acid metabolism, and choline metabolism (3-hydroxyisovalerate, 2-hydroxyisobutyrate, choline, N-methyl-2-pyridone-5-carboxamide, and N-methyl-nicotinamide). However, the researchers did not adjust for the major confounding factor of BMI, making it unclear whether differences in the BMI between cases and controls contributed to the metabolite differences observed. Graça et al. (2012) used UPLC-MS to investigate urine in the second trimester and discovered that choline was found to be higher in cases, a finding that confirmed the results of Diaz et al. (2011) who also identified choline as a discriminatory metabolite in the second trimester of cases. Sasche et al. (2012) investigated the urine metabolome at two different timepoints – week 8-20 (cohort number = 667) and week 26-30 (cohort number = 671); they found that only three metabolites were significantly different between cases and controls at both time points (citrate, and two unidentified peaks – all were significantly higher in GDM cases). Dudzik et al. (2014) performed a small case-control study (20 GDM cases, 20 controls) in the second trimester, analysing both plasma and urine using a multi-platform approach (LC-QTOF/MS, GC-Q/MS, and CE-TOF/MS). Compared to the reasonable number of plasma metabolites found to be significantly different between cases and controls, only two metabolites were significantly different in the maternal urine; histidine was higher in cases, and carnitine was lower. Law et al. (2017a) investigated changes in the urine metabolome as pregnancy progressed (urine collected at first, second, and third trimesters) using a case-control study (27 cases, 34

controls) of participants with and without diagnosed GDM in China. The authors found metabolites associated with tryptophan metabolism and purine nucleosides to be in significantly higher levels in GDM cases when compared to controls, throughout the course of the pregnancy.

The first investigation of the plasma metabolome was conducted by Diaz et al. (2011) using samples procured in the second trimester, and analysed using NMR. Betaine and trimethylamine-N-oxide (TMAO) were found to be significantly lower in cases when compared with controls. Betaine has been related to choline metabolism and this finding is thus consistent with the dysregulated choline metabolism found in the urine metabolome. De Seymour et al. (2014) were the first to analyse the maternal serum metabolome related to GDM and found one metabolite that was significantly higher in cases - itaconic acid. The authors hypothesised that the higher relative abundance of itaconic acid in the cases may have been associated with meta-inflammation above that expected with obesity/increased adiposity (as case and control groups were matched for BMI). Dudzik et al. (2014) performed a small case-control study (20 GDM cases, 20 controls) in the second trimester, analysing plasma using a multi-platform approach (LC-QTOF/MS, GC-Q/MS, and CE-TOF/MS). The author's found that lysophospholipids were the group of compounds most significantly affected by GDM; LPE(20:1) and LPE (20:2) were significantly lower in GDM cases when compared to controls, with the highest area under the receiver operating characteristic curve Liu et al. (2015) used UPLC-MS to explore the third trimester serum metabolome of 12 women who were diagnosed with GDM compared to 10 healthy controls. The serum samples were collected in the third trimester, when participants were close to delivery, and as such, the differences observed in the serum metabolome in Table 1.6 are highly confounded by any strategies to manage GDM following diagnosis, which are not detailed in the publication. Liu et al. (2015) were not the only group to investigate the metabolome post-diagnosis: Pinto et al. (2015) found that in the maternal plasma obtained post-diagnosis, cases demonstrated lower levels of glutamine, creatinine, betaine, proline, methanol and glucose. Lower levels of glucose in GDM cases was an unusual finding and may reflect some of the limitations of this particular study. Pinto et al. stated that the unusual glucose finding "may be explained by the fact that subjects were informed of GDM diagnosis approximately 2 weeks prior to sample collection having, probably, privately adopted lower sugar diets prior to the formal layout of the GDM management scheme" (p. 2704). Dudzik et al. (2017) performed an untargeted longitudinal study of the maternal plasma metabolome related to GDM using GC-MS

analysis of 24 GDM cases and 24 healthy controls. The authors found three potential biomarkers of GDM in the second trimester plasma that showed good predictive ability, with AUC of 0.9, 0.87, and 0.9, respectively for 2-hydroxybutyrate, 3-hydroxybutyrate, and stearic acid. They also showed that the metabolomic differences observed in second trimester plasma were attenuated following delivery in women who experienced normal glycaemic control. Law et al. (2017) analysed the maternal plasma metabolome longitudinally; analysing the metabolome of 61 pregnant women (27 GDM cases, 34 controls) whose plasma was collected in the first, second, and third trimesters. Law et al. found a large number of significant differences between the GDM cases and controls throughout the various timepoints, the most significant of which were the polyunsaturated or chemically modified phospholipids which were significantly lower in women with GDM. This finding was consistent throughout the progression of pregnancy. However, Law et al. did not adjust for multiple comparison testing and as such, a large number of their significant findings are likely to be false positive findings.

A small pilot study has conducted the first investigation into the use of maternal hair for differentiation of GDM cases from controls. He et al. (2015) performed the first investigation into the hair metabolome related to GDM, using GC-MS (number of cases = 47; number of controls = 47). They discovered one metabolite that differed significantly between cases and controls – adipic acid – which has been hypothesised to be related to altered fatty acid metabolism in response to lipid peroxidation in gestational diabetes. As this was the first study to use the hair metabolome, further studies are required to validate these findings in a larger cohort.

In a systematic review of metabolite profiling in GDM, published in *Diabetologia* (Huynh, Xiong, & Bentley-Lewis 2014), the authors criticised the lack of consistent findings. This inconsistency is also apparent in the assessment of untargeted metabolomics studies. The use of different analytical platforms, metabolite extraction protocols, data processing, and most importantly, specimen type, limited the ability to distinguish and validate a distinct metabolomic profile of GDM. Despite the disparate experimental designs, there does appear to be some consistency in the report of altered choline metabolism in women who developed GDM, compared with controls. However, these findings are yet to be validated in large cohort studies and have therefore not yet led to the identification of any robust predictive biomarkers.

Most of the studies investigating the GDM metabolome have relied on small sample numbers. It is apparent that future research in this area would benefit from studying the metabolome related to GDM in large, well-characterised pregnancy cohorts, taking into account potential confounding variables such as BMI, age, and ethnicity when comparing case-control differences. Large cohort studies have the statistical power necessary to detect meaningful significant differences in a human population where inter-individual variation can be significant. In addition to utilising metabolomics for exploring biomarkers of GDM development, the metabolome provides useful information about the metabolic dysregulation that occurs in GDM and can therefore inform researchers of appropriate metabolic pathways to target for the development of interventions and prevention strategies. One such way that this can be achieved is by linking the metabolome in GDM to a modifiable lifestyle factor such as the maternal diet.

1.4.3.1 The link between the diet and the metabolome.

The human diet has a strong influence on metabolism, health and disease. Consequently, dietary intake has an observable effect on the human metabolome, and now there is an entire field of research dedicated to understanding the effects of dietary consumption on the metabolome – the food metabolome. The food metabolome has been defined as "the part of the human metabolome directly derived from the digestion and biotransformation of foods and their constituents" (Scalbert et al., 2014: p. 1286). There is a growing body of evidence that has identified metabolomics-derived dietary biomarkers (O'Gorman, Gibbons, & Brennan, 2013), from food items such as citrus fruit, to walnuts, coffee, and cocoa, and even some dietary patterns (Garcia-Aloy et al., 2014; Garcia-Aloy et al., 2015; Garcia-Perez et al., 2017; Heinzmann et al., 2010; Rothwell et al., 2014; Vázquez-Fresno et al., 2015). However, food metabolome research has predominantly revolved around assessing the effect of interventions on the metabolome, or dietary consumption in subjects who are under controlled environments. Far fewer studies have been performed to investigate whether dietary intake can be differentiated using the metabolome of free-living subjects. It is unknown whether the findings in controlled situations can be replicated in free-living subjects who have a much greater variability in their day-to-day dietary intake. Table 1.7 details some of the studies conducted thus far that have differentiated the diet of free-living subjects using metabolomics.

Table 1.7

Studies Linking Dietary Intake of Free-Living Participants to Changes in the Human Metabolome

Authors & date	Participants	Sample type	Analytical platform	Dietary assessment	Findings
O'Sullivan et al. (2011)	125 participants	Urine	NMR	3-day food diaries. Dietary patterns determined using cluster analysis	Cluster 1 higher energy contribution from wholemeal bread, whole milk, fish, confectionary, and ice cream and desserts and a lower contribution from low-energy beverages. Cluster 3 higher contributions of white bread, sugars and preserves, butter and spreads, red meat, red-meat dishes, meat products, and alcohol and a lower contribution from vegetables. TMAO, glycine, acetylcarnitine, phenylacetylglutamine, acetoacetate, and nn-dimethylglycine were the metabolites that differentiated cluster 1 from cluster 3. Urine levels of phenylacetylglutamine were positively associated with vegetable intake. Urine levels of O-acetylcarnitine were positively associated with red meat intake.

Authors & date	Participants	Sample type	Analytical platform	Dietary assessment	Findings
De Filippis et al. (2016)	153 participants (51 vegetarians, 51 vegans and 51 omnivores)	Faeces Urine	GC-MS NMR	7-day weighed food diaries.	Urinary TMAO levels significantly lower in vegetarian and vegan participants in comparison to omnivores. Higher scores on the Mediterranean diet index (MDI) were associated with lower levels of TMAO. Acetic acid, propanoic acid and butanoic acid from the faecal samples were associated with the consumption of fruit, vegetables, legumes and fibre. Valerate and caproate levels were associated with higher intake of protein-rich animal foods and fat intake. Propanoate and acetate were positively associated with MDI scores in omnivores.
Xu et al. (2010)	161 participants (81 omnivores, 80 lactovegetarians)	Urine	NMR	None – just recruitment criteria in regards to which dietary-restricted group they belonged too.	Compared to omnivores, the lactovegetarians had significantly lower levels of creatinine, taurine and TMAO in their urine.

Authors & date	Participants	Sample type	Analytical platform	Dietary assessment	Findings
Lloyd et al. (2011)	23 participants	Urine	FIE-MS	Validated FFQ to classify citrus consumers.	Urine levels of proline betaine were positively associated with habitual citrus consumption.
Floegel et al. (2013)	2380 participants	Serum	FIA-MS/MS	148-item FFQ used to derive dietary patterns using reduced rank regression.	A dietary pattern high in butter and low in margarine was positively associated with higher saturated acylcarnitines, such as C9, C16 and C18, and lower unsaturated acylcarnitines (C8:1, C18:2 and C14:2). A dietary pattern low in fish, vegetables and whole-grain bread was associated with higher levels of C14, C2 and C18 acylcarnitines. A dietary pattern high in potatoes, cornflakes, dairy products, raw vegetables and desserts and low in soup and beverages was associated with amino acid levels, especially methionine. A dietary pattern high in canned fruit, fried potatoes and legumes and low in water, low-fat cheese and fish was associated with higher levels of serine. A dietary pattern high in red meat, non-whole-grain bread and fish and low in whole-grain bread and tea was associated with higher levels of hexose.

Authors & date	Participants	Sample type	Analytical platform	Dietary assessment	Findings
					A dietary pattern high in fish and poultry and low in sweet foods, margarine, tea and whole-grain bread was associated with levels of diacyl-phosphatidylcholines (particularly saturated and polyunsaturated). A dietary pattern low in fish, whole-grain bread and tea, and high in sauce and butter was associated with the monounsaturated diacyl-phosphatidylcholines. A dietary pattern high in butter, red meat and high-fat dairy products and low in margarine and whole-grain bread was associated with acyl-alkyl-phosphatidylcholines. A dietary pattern high in meats and margarine and low in butter and sweet foods was associated with the unsaturated acyl-alkyl-phosphatidylcholines. A dietary pattern high in butter, high-fat-dairy and sweet bread spreads and low in margarine, low-fat cheese and pasta/rice was associated with levels of lyso-phosphatidylcholines. A dietary pattern high in margarine, non-whole-grain bread, meat and coffee and low in butter, pasta/rice and tea was associated with lyso-

Authors & date	Participants	Sample type	Analytical platform	Dietary assessment	Findings
					phosphatidylcholines C20:4 and C18:2. A dietary pattern high in butter, garlic and coffee and low in margarine, fresh fruit and soup was associated with levels of sphingomyelins (particularly the hydroxyl-sphingomyelins). A dietary pattern low in butter, sweet foods, high-fat dairy, fruits and whole-grain bread was associated with the sphingomyelins without the hydroxy group.
Menni et al. (2013)	1003 females (75 monozygotic twin pairs, 228 were dizygotic twin pairs and 397 singletons)	Serum	FIA-MS/MS	131-item FFQ 5 dietary patterns derived using PCA.	Fruit and vegetable pattern (frequent intake of fruit, allium and cruciferous vegetables; low intake of fried potatoes) was associated with higher levels of sphingomyeline C26:1 and phosphatidylcholine diacyl C38:6. The dieting dietary pattern (frequent intakes of low-fat dairy products, low-sugar soda; low intake of butter and sweet baked products) was associated with higher levels of nonacylcarnitine (C9). Coffee intake was associated with lower levels of decenoylcarnitine (C10:1). Garlic intake was associated with lower levels of octenoylcarnitine (C8:1) and glutarylcarinitine (C5–DC/C6–0H).

Authors & date	Participants	Sample type	Analytical platform	Dietary assessment	Findings
Bouchard-Mercier et al. (2013)	37 participants	Plasma	FIA-MS/MS	91-item validated FFQ. 27 food groups used to generate dietary patterns using PCA.	After adjustments for age, BMI, and gender: The Western dietary pattern (high intakes of refined grain products, desserts, sweets and processed meats) was negatively associated with principal component 1 of the metabolites (predominantly medium- to long-chain acyl-carnitines). The Western dietary pattern was positively associated with principal component 2 of the metabolites (predominantly amino acids and short-chain acyl-carnitines).
Schmidt et al. (2015)	397 men (98 vegans, 98 meat eaters, 98 fish eaters [who do not eat meat but do eat fish], and 98 vegetarians)	Plasma	LC-MS/MS	None – grouped according to dietary group at recruitment.	Metabolites were grouped into four principal components: The first component was high in some glycerophospholipids (especially PC aa C32:1, PC aa C36:5, PC aa C36:6, and PC ae C34:0) and some sphingolipids (especially SM(OH) 14:1). Principal component 2 was high in three acylcarnitines (C-16, 18:1, and C18:2) and low in arginine and hexose. The third component was high in glycerophospholipids (especially PC aa 32:1, PC aa C34:4, and PC aa C40:5) and low in some sphingolipids (especially SM(OH) 16:1). The fourth component was high in some glycerophospholipids (especially PC ae C44:6).

Authors & date	Participants	Sample type	Analytical platform	Dietary assessment	Findings
					Men who ate meat had the highest scores on PC1 and vegans had the lowest score on PC1, after adjustment for potential confounders. Men who followed a vegan diet had higher scores on PC3 and PC4, after adjustment for potential confounders.
Edmands et al. (2015)	481 participants	Urine	UPLC-MS	FFQ to determine 6 polyphenol-rich foods: coffee, tea, red wine, citrus fruit, apples, pears, chocolate.	817 features were significantly associated with habitual coffee consumption, 499 features with red wine consumption, 7 features with citrus fruit consumption, 36 features with tea consumption, and 10 features with apple and pear consumption.
Gibbons et al. (2015)	565 participants	Urine	NMR	4-day semi-weighed food record.	Consumption of sugar-sweetened beverages was associated with higher levels of citrulline, formate, isocitrate, and taurine.
Garcia-Aloy et al. (2015)	275 participants (56 non-consumers of bread, 48 white-	Urine	HPLC–MS	Validated semi-quantitative 137-item FFQ.	Higher levels of HMBOA and riboflavin could discriminate non-bread consumers from bread consumers. Higher levels of 2-hydroxy-N-(2-hydroxyphenyl) acetamide, enterolactone glucuronide, 5-(3,5-dihydroxyphenyl) pentanoic acid,

Authors & date	Participants	Sample type	Analytical platform	Dietary assessment	Findings
	bread consumers, 51 wholegrain-bread consumers)				dihydroferulic acid, and pyrraline could discriminate wholegrain-bread intake from white-bread intake and non-consumers.
Rothwell et al. (2014)	20 high coffee consumers (median intake of 290 mL/day) 19 low coffee consumers (0mL/day)	Urine	UPLC-QTof-MS	Validated FFQ	119 ions in positive mode and 13 ions in negative mode were significantly different between high and low coffee consumers. Atractyligenin glucuronide and diketopiperazine cyclo(isoleucyl-prolyl) were two novel biomarkers of coffee intake that were identified by this study.

The studies detailed in Table 1.7 have been successful in differentiating dietary habits using the metabolome of free-living subjects, and include identification of metabolomic markers of individual food group consumption (Gibbons et al., 2015; Menni et al., 2013; Rothwell et al., 2014), as well as habitual dietary-pattern consumption (Bouchard-Mercier, Rudkowska, Lemieux, Couture, & Vohl, 2013; O'Sullivan, Gibney, & Brennan, 2011). With convincing evidence that the diet has an effect on the metabolome and overall metabolism, it is reasonable to hypothesise that the influence of diet on disease could be better understood by exploring the effect of the diet on the metabolome, and combining it with information on how the metabolome is related to the disease of interest.

There are two main ways that the metabolome can be used to investigate the relationships between diet and disease. Firstly, identifying metabolites that are both markers of food-group or dietary-pattern intake, and are implicated in a disease trajectory, allows for identification of dietary biomarkers associated with disease. Dietary biomarkers of disease can link dysregulated metabolism associated with disease to a modifiable lifestyle factor, providing a potential avenue to explore for intervention. Secondly, understanding the link between metabolic pathways that are related to dietary intake can assist with understanding and hypothesis generation of *how* diet is linked to diseases. By exploring *how* diet is linked to disease from a metabolomics stance, researchers can gain better insight into metabolic pathways that could be targeted through dietary interventions to prevent or manage disease.

1.4.5 Identification of research need.

Studies to date that have investigated the metabolome and its association with GDM have thus far failed to detect any clinically useful biomarkers for early identification of women at risk of GDM development. The studies conducted have predominantly used small sample sizes, and the field is likely to benefit from large, well-characterised cohort studies in the search for biomarkers. In addition, future studies should account for confounding variables to differentiate the effects of GDM on the metabolome from the effects of demographic risk factors that co-exist, such as advanced maternal age, high maternal BMI, and ethnicity. With some of the highest rates of GDM occurring in Asian countries, it is also paramount that more studies are conducted in this at-risk population, to advance identification and prevention strategies that can be appropriately translated to these high-risk populations.

As discussed in Section 1.3, there are limitations to dietary investigations and determining the association of diet with disease. In the case of GDM, a large number of studies have

investigated the association between maternal diet and GDM. However, many have taken a single-nutrient approach and limited studies to date have sought to understand *how* and *why* the observed dietary component is related to GDM development, in terms of its effect on metabolism. Understanding *how* and *why* a dietary component is related to GDM is crucial for the development of successful GDM prevention strategies. The metabolome is one way in which this could be achieved.

1.5 Summary

Gestational diabetes mellitus is a pregnancy disorder with adverse consequences for both the mother and her offspring. The growing prevalence of GDM signifies a need to develop strategies to prevent its development and associated consequences. With a large proportion of unplanned pregnancies, the earliest feasible stage of intervention is during pregnancy. However, to intervene, women who are at risk of GDM development first need to be identified, to target interventions. The current clinical risk-factors used for identification fail to identify a significant proportion of women who go on to develop GDM. Many women are not aware of their GDM risk until diagnosis in late pregnancy (usually between the 24th and 28th weeks of gestation). It is evident that an early pregnancy screening tool, to predict GDM development, is warranted. Metabolomics is an analytical approach that can be employed in the search for biomarkers of GDM. Metabolomics is the real-time assessment of metabolism which is influenced by environment and lifestyle, therefore reflecting the phenotype of an individual. The metabolome's reflection of the phenotype offers an opportunity to explore modifiable environmental and lifestyle factors that could be contributing to GDM and its consequences.

To prevent the development of GDM after identification, an intervention needs to be established. Without an intervention strategy, a biomarker becomes less useful clinically, as it is difficult to justify the expense of a screening tool if nothing can be done to change the trajectory of the disease, or reduce its adverse outcomes. The maternal diet poses a safe, modifiable lifestyle factor that could be explored for the prevention of GDM. Despite the large range of studies conducted, there is little agreement as to the best dietary intervention to prevent GDM. This is likely due to a number of limitations in the studies conducted so far, in conjunction with the fundamental limitations of typically employed observational dietary assessments. Understanding *how* the diet is associated with GDM will help in the development of a dietary strategy for the prevention of GDM and its consequences.

Metabolomics is an analytical tool that is capable of shedding light on the effects of the maternal diet on metabolism, and therefore relating these effects to the dysregulated metabolism in GDM.

1.6 Aim and Objectives

1.6.1 Aim.

The aim of this thesis is to identify associations in order to generate hypotheses relating to the relationship between maternal diet and GDM, by exploring the effect of both diet and disease on the maternal metabolome.

1.6.2 Objectives.

1. To analyse 24-hr recalls of maternal dietary intake, identify dietary patterns and investigate associations with GDM;
2. To analyse the maternal plasma metabolome using GC-MS, and assess whether the metabolome can differentiate women who developed GDM from those who did not;
3. To analyse the maternal hair metabolome using GC-MS, and assess whether the metabolome can differentiate women who developed GDM from those who did not;
4. To apply statistical analyses to look for connections between the 24-hr dietary recall data (dietary patterns and food groups), GDM, and the maternal plasma and hair metabolome.

Chapter Two: Study Design and Framework

2.1 Growing Up in Singapore Towards Healthy Outcomes (GUSTO) Study

The GUSTO study is a multi-ethnic Asian prospective pregnancy cohort based in Singapore. The study was established in June, 2009, and the pregnancy biobanking was completed by September, 2010. Participants were from three main ethnic groups in Singapore: Malay, Indian, and Chinese. Eligibility criteria for participation in the study specified that the parents of participants were from the same ethnic group. Recruitment occurred during the participants' first trimester of pregnancy and women were recruited at both Kandang Kerbau Women's and Children's Hospital (KKH) and National University Hospital (NUH). A total of 1247 women were recruited into the GUSTO study. The study was approved by the National Health Care Group Domain Specific Review Board (reference D/09/021) and the Sing Health Centralized Institutional Review Board (reference 2009/280/D). Research was conducted according to the Declaration of Helsinki and all participants gave their written informed consent at recruitment.

The four studies in this thesis were conducted using data and biological samples from a sub-set of women recruited to the GUSTO cohort. Each chapter outlines the eligibility criteria for inclusion into that specific study. Figure 2.1 shows the number of participants used in each chapter, and how the participants were interrelated in the studies described in Chapters 4 and 6. The reason for selecting the GUSTO cohort to perform the investigation of the relationship between maternal diet and GDM using a metabolomics approach, is that at the time of commencing this project, GUSTO was the only accessible cohort that had obtained both dietary data and biological samples during pregnancy, while also performing universal screening for GDM using OGTT. The GUSTO study also consisted of an Asian participant group – a population who have a high prevalence of GDM, and therefore require ongoing research into preventions that can translate to this at-risk population.

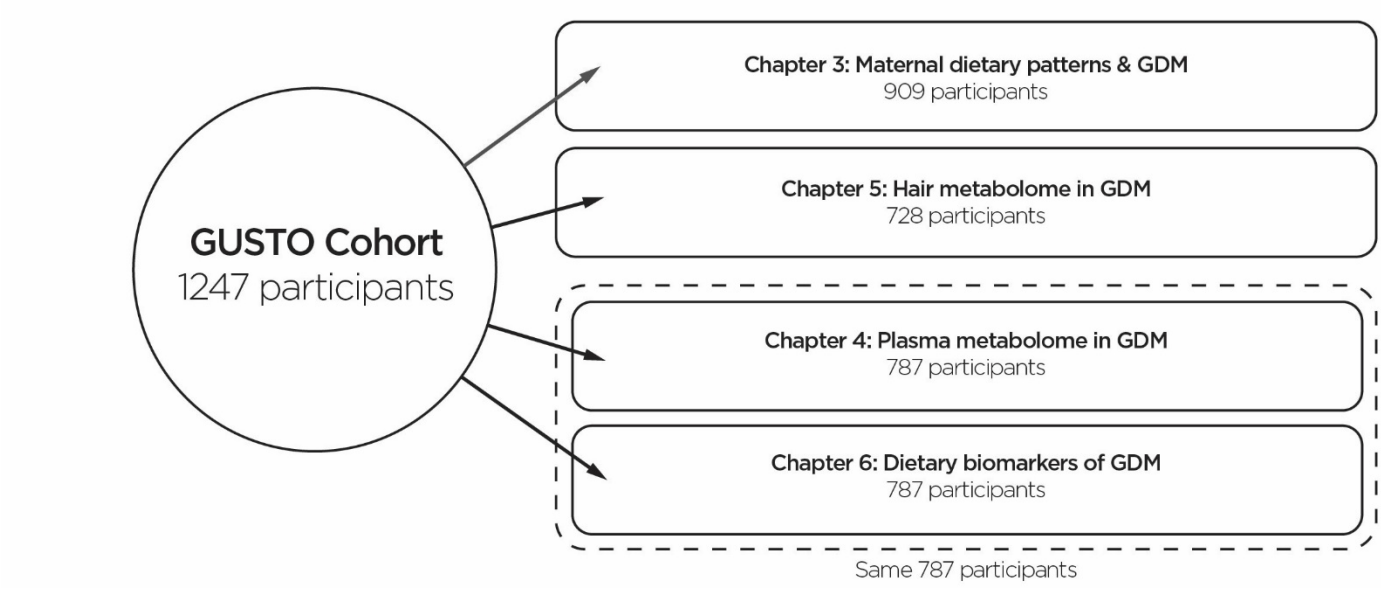


Figure 2.1. Participants from the GUSTO cohort analysed in each chapter of the thesis.

2.2 Study Design Framework

The following thesis chapters comprise four studies conducted using the GUSTO study participants. Each study provided information to assist with the overall aim of the thesis which was to generate hypotheses on the relationship between maternal diet and GDM, by exploring the effect of both diet and disease on the metabolome:

1) Chapter 3 investigates the relationship between maternal diet and GDM,
2) Chapter 4 investigates the plasma metabolome associated with GDM,
3) Chapter 5 investigates the hair metabolome associated with GDM,
4) Chapter 6 investigates the relationship between the maternal diet and the GDM-related metabolome.

The framework of this thesis is that the maternal dietary association to GDM was first defined using a subjective and observational approach, where self-reported dietary recalls were statistically analysed for their association to GDM, an approach which is commonly used in dietary association studies. The plasma and hair metabolomes were then analysed to determine how they differed between women who did and did not develop GDM. Finally, the metabolomics and dietary datasets were combined to explore whether the associations between diet and GDM could be explained in the metabolome, providing an objective and

mechanistic insight into how the diet was affecting metabolism, and, in turn, was related to GDM.

This framework is visually represented in Figure 2.2

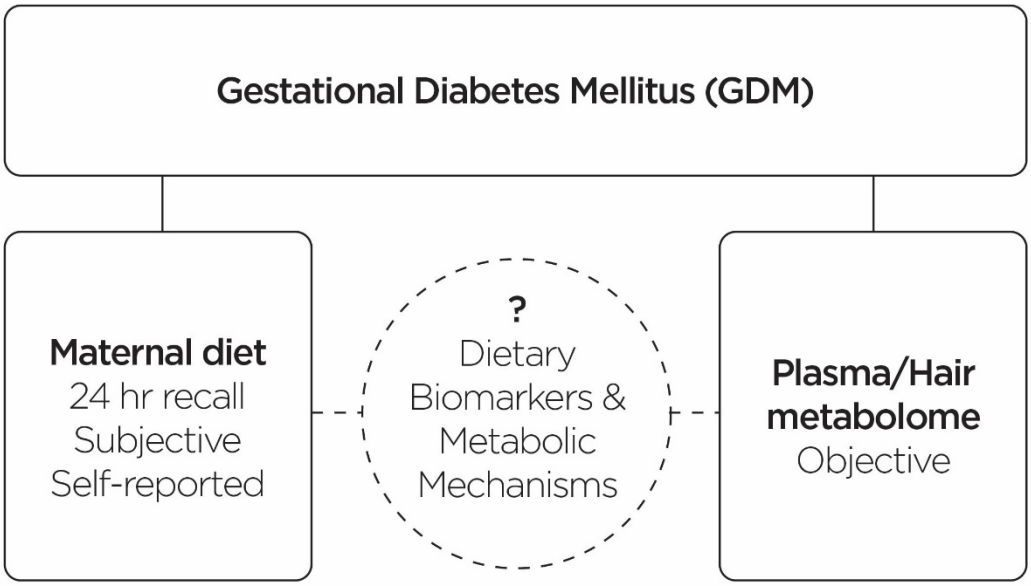


Figure 2.2. Framework underpinning the four studies conducted in this thesis to meet the overall aim.

Chapter Three: Maternal Dietary Patterns and Gestational Diabetes Mellitus

3.1 Introduction

Despite the abundance of research demonstrating an association between maternal diet and GDM (see Appendix 1 for a summary of the most recent literature), no unequivocal link has yet been substantiated. The majority of studies conducted thus far have focused on individual nutrients. Addressing nutrients as individual, independent entities, rather than adopting a holistic approach, is a possible reason for these discordant findings.

3.1.1 Dietary pattern analysis.

Dietary pattern analysis is an integrated approach to understanding the dietary consumption of an individual or group that takes into account the combination of foods and nutrients consumed. Dietary pattern analysis is a data-driven, bottom-up scientific approach to dietary investigations where rather than using an a priori method to investigate associations between diet and disease, statistical techniques are applied to let the data inform the researcher of associations. The two most commonly employed methods for dietary pattern discovery are factor analysis and cluster analysis (Hu, 2002). A description of the statistical approaches used to derive dietary patterns can be found in the literature review section of this thesis (Section 1.3.3). Previous studies exploring maternal dietary patterns in relation to GDM are summarised in Table 1.4.

3.1.2 Ethnicity.

The majority of dietary pattern studies were performed in Western populations. Only one study to date has conducted dietary pattern investigations related to GDM in an Asian population (He et al., 2015). However, the rates of raised blood glucose in Asia are some of the highest worldwide (International Food Policy Research Institute, 2016), and the prevalence of GDM in South-East Asia is > 20% and growing (IDF, 2015). It is clear that this is an at-risk population that requires intervention. Therefore, understanding the role that the local diet plays in GDM development in this population is very important. The dietary patterns generated by He et al. (2015) were distinctly different to those found in previous studies conducted in the Western populations, demonstrating the uniqueness of typical dietary consumption in Asia. This uniqueness implies a need for further studies in non-Western

populations, and demands a cautious approach to the translation of dietary findings across continents and cultures.

To address this gap in the research, dietary recall information collected from the GUSTO study cohort – a large multi-ethnic Asian cohort, was analysed to derive maternal dietary patterns and explore their relationship with GDM.

3.2 Methods

3.2.1 Participants.

The investigation of the maternal diet was performed on a sub-set of women from GUSTO. Details of the GUSTO study cohort are described in Chapter 2. The GUSTO participants were eligible for this dietary pattern study if they met the following criteria:

a) Underwent an oral glucose tolerance test (OGTT) at 26-28 weeks' gestation;

b) Completed a 24-hr dietary recall at the time of the OGTT;

c) Had complete age, body mass index (BMI), and ethnicity data (as these are three major clinical predictors of GDM).

Figure 3.1 displays participant exclusion criteria used for this study. Maternal ethnicity, age, education level, family history of diabetes, previous history of GDM, and monthly household income was obtained at recruitment, and BMI, body fat estimates (sum of five skinfolds), cigarette smoking and alcohol consumption during pregnancy was recorded at the time of the OGTT (26th - 28th weeks of gestation). To calculate pregnancy BMI, maternal weight was measured using digital body weight scales (SECA model 803) and standing heights were measured with a stadiometer (SECA model 213). BMI was also collected at time of recruitment but was not used in analyses as it was self-reported and had a high rate of non-disclosure.

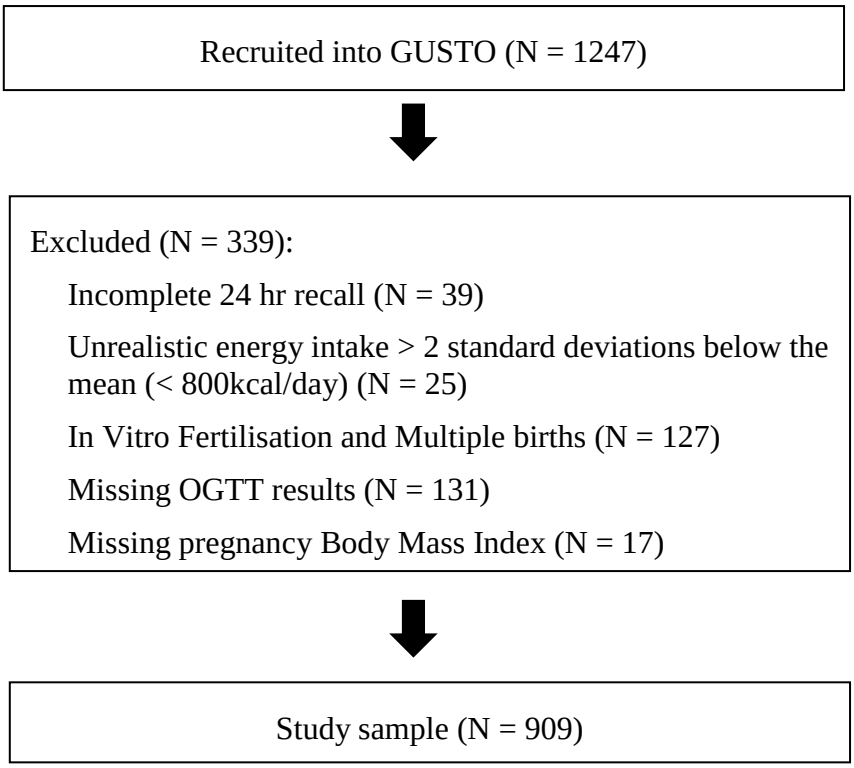


Figure 3.1. Participant exclusion flowchart shows the number of participants excluded as a result of each exclusion criteria.

3.2.2 Dietary assessment.

Maternal diet was assessed using a 24-hr dietary recall collected at the time of the OGTT (between 26-28 weeks of gestation). The 24-hr dietary recalls were interviewer-administered using the five-stage multiple-pass recall interviewing technique (Conway, Ingwersen, Vinyard, & Moshfegh, 2003). This technique is used to improve participants' recall. Information from dietary recalls was entered into electronic spreadsheets (Microsoft Excel) by research staff at the National University of Singapore (NUS). For validation purposes, 3-day food diaries were also collected from a small sub-set of participants (N = 208) in the week following the OGTT. Participants were trained by clinical staff on how to correctly complete the 3-day food diaries. Nutritional composition of the maternal diet was determined using a locally relevant dietary analysis software (Dietplan 7, Forestfield Software). Foods reported in the 24-hr dietary recalls were assigned to one of 68 food groups, which were grouped in collaboration with a panel of nutrition experts within Dr Mary Chong's research group (Singapore), according to similar nutrient composition (Table 3.1).

3.2.3 Statistical analysis.

Statistical analysis was performed using SPSS version 23.0.

Missing values for discrete demographic variables were imputed using a multiple imputation procedure (Rubin, 1996) in SPSS. Imputed variables included birth order (N = 10), education level (N = 11), monthly household income (N = 58), family history of diabetes (N = 33), previous history of GDM (N = 33), smoking (N = 1) and alcohol consumption during pregnancy (N = 21). An assumption of the imputation procedure is that the values are missing at random. However, it is possible that some of the missing values may have been due to a social desirability response bias, particularly in relation to smoking and alcohol consumption during pregnancy. It is also possible that sensitive questions such as household income and education level may not have been disclosed for similar reasons. However, without proof of such events, this is only speculation and therefore, for this study, it has been assumed that the information is missing at random.

Exploratory principal components factor analysis, a well-established approach for identifying dietary patterns (He et al., 2015; Newby, Muller, Hallfrisch, Andres, & Tucker, 2004; Osler, Heitmann, Gerdes, Jùrgensen, & Schroll, 2001; C. Zhang, Schulze, Solomon, & Hu, 2006), was performed on standardised values of participant food intake from the 68 food groups, to identify the dietary patterns in the cohort. Orthogonal (varimax) rotation was performed to ensure that the resulting factors were uncorrelated and therefore allow for the use of dietary pattern scores as predictors in downstream regression analyses. Ensuring factors were uncorrelated means that their contribution in regression analyses could be interpreted independently of one another. To define the number of dietary patterns to be used for further analysis, the following criteria were applied:

1) The number of components identified at the elbow of the scree plot of eigenvalues (> 1) following the factor analysis (Figure 3.2), and;

2) The number of components that produced interpretable dietary patterns, relevant to typical food consumption in Singapore.

Food groups with a loading factor > 0.25 or <-0.25 were used to describe the patterns, and the main food groups in each dietary pattern were used to name that pattern. Food groups and corresponding loading scores for each dietary pattern are listed in Table 3.2, and visually represented in Figure 3.2. Dietary pattern scores for each participant were calculated as the

sum of the standardised food group intake multiplied by the corresponding factor loading, for each pattern.

All foods reported in the 3-day food diaries were assigned to one of the 68 pre-defined food groups. An applied dietary pattern score was calculated for the dietary patterns significantly contributing to GDM, as determined by downstream analyses (Crozier, Robinson, Godfrey, Cooper, & Inskip, 2009). Applied scores were calculated by multiplying standardised food group intake from the 3-day food diaries with the factor loading of that food group in the dietary pattern derived from the 24-hr recalls, then taking the sum of the results from all 68 food groups. Pearson's product-moment correlation was performed between the applied scores and the scores from the 24-hr recalls.

Binary logistic regression analysis was performed to investigate the relationship between the dietary patterns and GDM. The regression analyses were adjusted for demographic variables found to be significantly different between GDM cases and controls in the cohort, therefore defined as confounding variables: maternal age, maternal BMI, previous history of GDM, ethnicity, and education level (as a proxy for household income level also, as both were highly correlated; Pearson's $R = 0.56$, $P < 0.001$). Macronutrient composition was compared between participants in the highest and lowest tertiles of dietary pattern scores for each of the dietary patterns that were found to be significantly associated with GDM. For all statistical analyses, a 2-tailed P-value cut-off of less than 0.05 was considered to be statistically significant.

3.2.4 Sensitivity analysis.

To assess the effect of imputation on downstream analyses, dietary patterns were generated and regression analyses performed using a dataset of participants who had complete information for all variables (N = 822), for comparison with the imputed dataset (N = 909).

Table 3.1

List of the 68 Food Groups

Food Groups		
Poultry (B)	Grains	Cheese
Red Meat (B)	Oats	Other Dairy
Red Meat (F)	Breakfast Cereals/Bars	Carbonated Drinks
Meat Products	Butter/Ghee	Sweetened Drinks
Meat Innards	Margarine/Peanut Butter	Soya Bean Drinks
Fish (B)	Salad Dressing	Coffee/Tea
Fish (F)	Cream-based Soup	Dessert Soup
Seafood	Soya Sauce-based Gravy	Local Sweet Snacks/Pastries/Biscuits
Fish/Seafood Products	Other Gravy	Local Savoury Snacks (B)
Eggs	Sweet Condiments	Local Savoury Snacks (F)
Vegetables	Chocolate	Ice cream
Starchy Vegetables	Noodles in Soup	Chips
Potatoes (F)	Flavoured Noodles	Blended Oil
Fresh Fruit	Pasta	MUFA/PUFA Oil
Fruit Juice	White Bread	Meat/Vegetable Soup
Processed Fruit	Wholemeal/Multigrain Bread	Tomato-based Gravy
Beancurd	Ethnic Bread	Cream-based Gravy
Legumes/Pulses	Bread with Toppings	Curry-based Gravy
Nuts/Seeds	Low-Fat Milk	Sweets
White Rice	Whole Milk	Sweet Spreads
Brown Rice	Formula Milk	Burger
Flavoured Rice	Milk-Based Drinks	Pizza
Porridge	Yogurt & Cultured Drinks	

F = Fried preparation or curry cooked in coconut

B = Boiled, steamed, grilled, roasted, baked, stir fried, braised, or stewed preparation

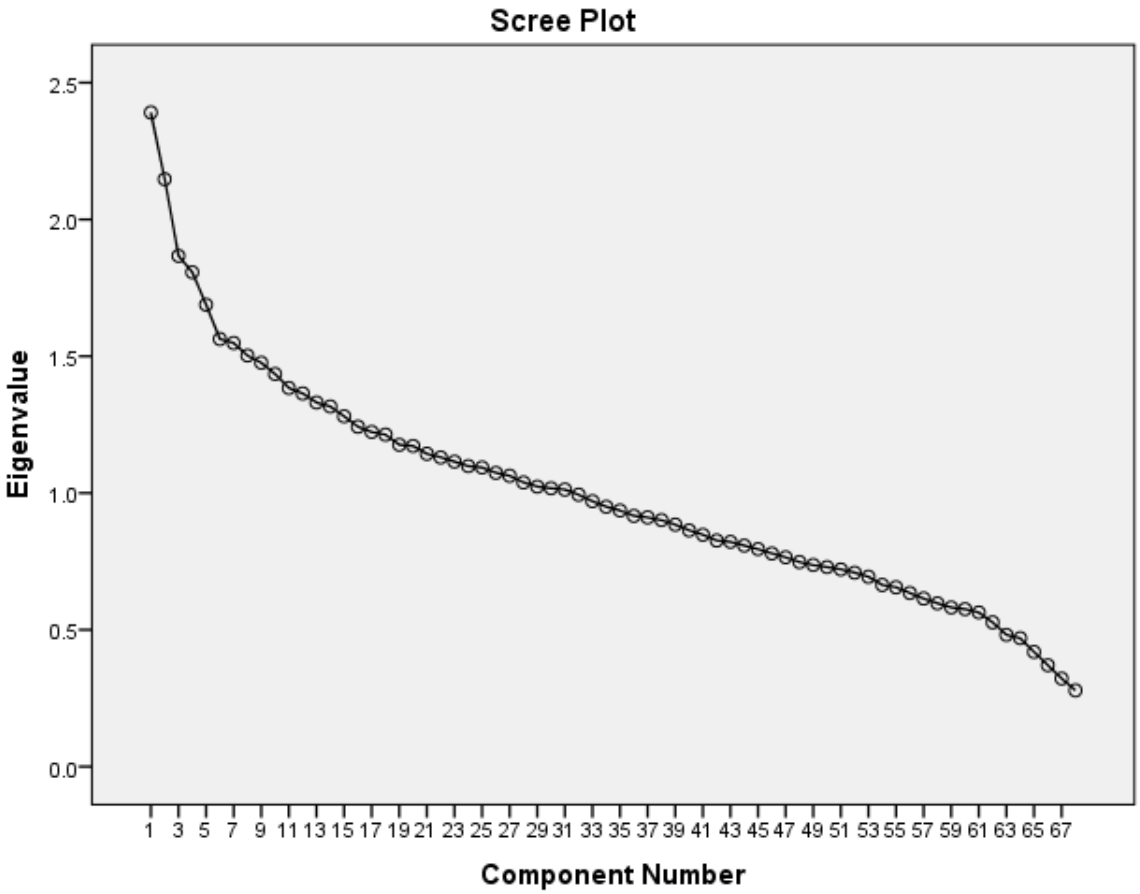


Figure 3.2. Scree plot of eigen values following factor analysis of 68 food groups (N=909)

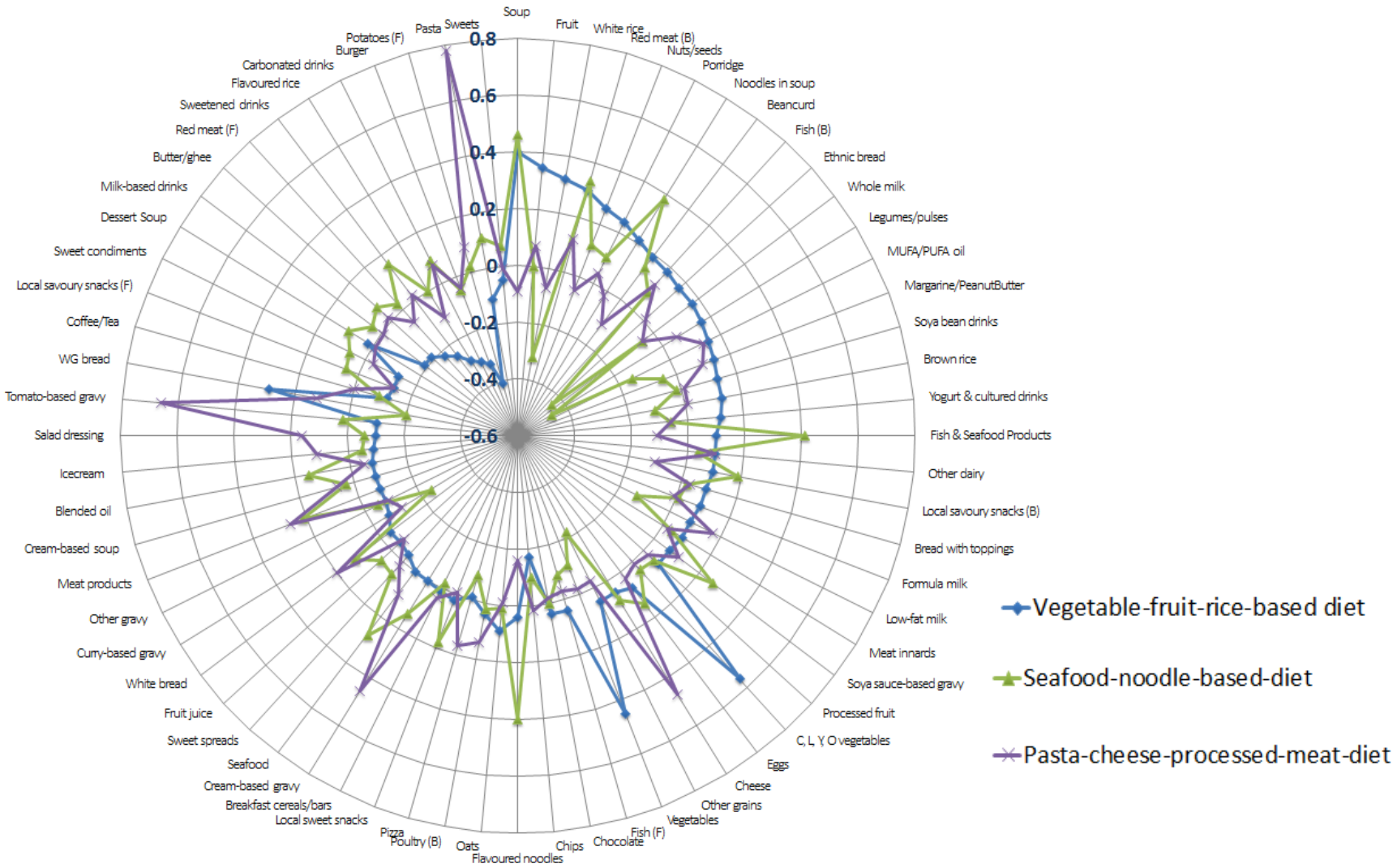


Figure 3.3. Radar plot to display the factor loadings of the 68 food groups in each of the three dietary patterns. The centre of the radar represents strong negative loading factors; the edge of the radar represents strong positive loadings.

Table 3.2

Three Dietary Patterns Identified from Exploratory Factor Analysis Using Varimax Rotation (N = 909)

Vegetable-fruit-rice-based Diet (3.5% variance explained)		**Seafood-noodle-based-diet (3.2% variance explained)**		**Pasta-cheese-processed-meat diet (2.7% variance explained)**	
Food Group	*Factor Loading Coefficient*	*Food Group*	*Factor Loading Coefficient*	*Food Group*	*Factor Loading Coefficient*
CLYO vegetables	0.54	Soup	0.54	Pasta	0.81
Other vegetables	0.45	Fish & seafood products	0.43	Tomato-based gravy	0.70
Fruit	0.40	Noodles in soup	0.41	Cream-based gravy	0.46
White rice	0.36	Flavoured noodles	0.39	Cheese	0.46
Soup	0.32	Red Meat (B)	0.35	Meat products	0.26
Wholemeal/ multi grain bread	0.29	Seafood	0.29		
Red meat (B)	0.26	Soya sauce-based gravy	0.26		
Red meat (F)	- 0.25	White rice	- 0.28		
Sweetened drinks	- 0.26	Curry-based gravy	- 0.28		
Flavoured rice	- 0.30	Legumes/Pulses	- 0.40		
Burger	- 0.32	Ethnic bread	- 0.42		
Carbonated drinks	- 0.33				
Potatoes (F)	- 0.41				

Extraction Method: Principal Component Analysis

CLYO vegetables = cruciferous, leafy, yellow, and orange vegetables

Rotation Method: Varimax with Kaiser Normalization

For display purposes, only food groups with a factor loading ≥ 0.25 or ≤ -0.25 are listed in the table.

F = Fried preparation or curry cooked in coconut

B = Boiled, steamed, grilled, roasted, baked, stir fried, braised, or stewed preparation

3.3 Results

3.3.1 Participant characteristics.

A total of 909 women were included in the study, and GDM was diagnosed in 17.6% (N = 160) of participants, according to the 1997 World Health Organisation (WHO) diagnostic criteria of fasting plasma glucose ≥ 7.0 mmol/l OR 2-hr plasma glucose ≥ 7.8 mmol/l following a 75 g glucose load (Alberti & Zimmet, 1998). The WHO criteria were used to diagnose GDM in this thesis, as these criteria allowed for identification of the highest risk group and had a higher cut-off criteria than other classification methods, as well as maintaining consistency with previous studies published from the GUSTO cohort (Chen et al., 2016; Chong et al., 2015; L. Li, Cheung, Ikram, Saw, & Wong, 2013). A principal components analysis plot was produced using the dietary data from all participants, in order to identify outliers (Figure 3.3). The first two components explained a very small amount of variation (3.5% and 3.2%) and there were no obvious outliers, therefore, no participants were removed from analysis.

Participant characteristics are detailed in Table 3.3. Participants were predominantly of Chinese ethnicity (56.7%), followed by Malay, (25.7%) and Indian (17.6%). Just under half of the participants (N=380, 41.8%) were nulliparous at recruitment, the mean age (± standard deviation) of participants was 31 (± 5) years, and the median BMI (lower quartile, upper quartile) was 25.3 (23.0, 28.2) kg/m^2. Participants diagnosed with GDM had a significantly higher age and BMI compared to those not diagnosed with GDM, were more likely to report GDM in a previous pregnancy, and reported a higher level of education and monthly household income level (Table 3.3).

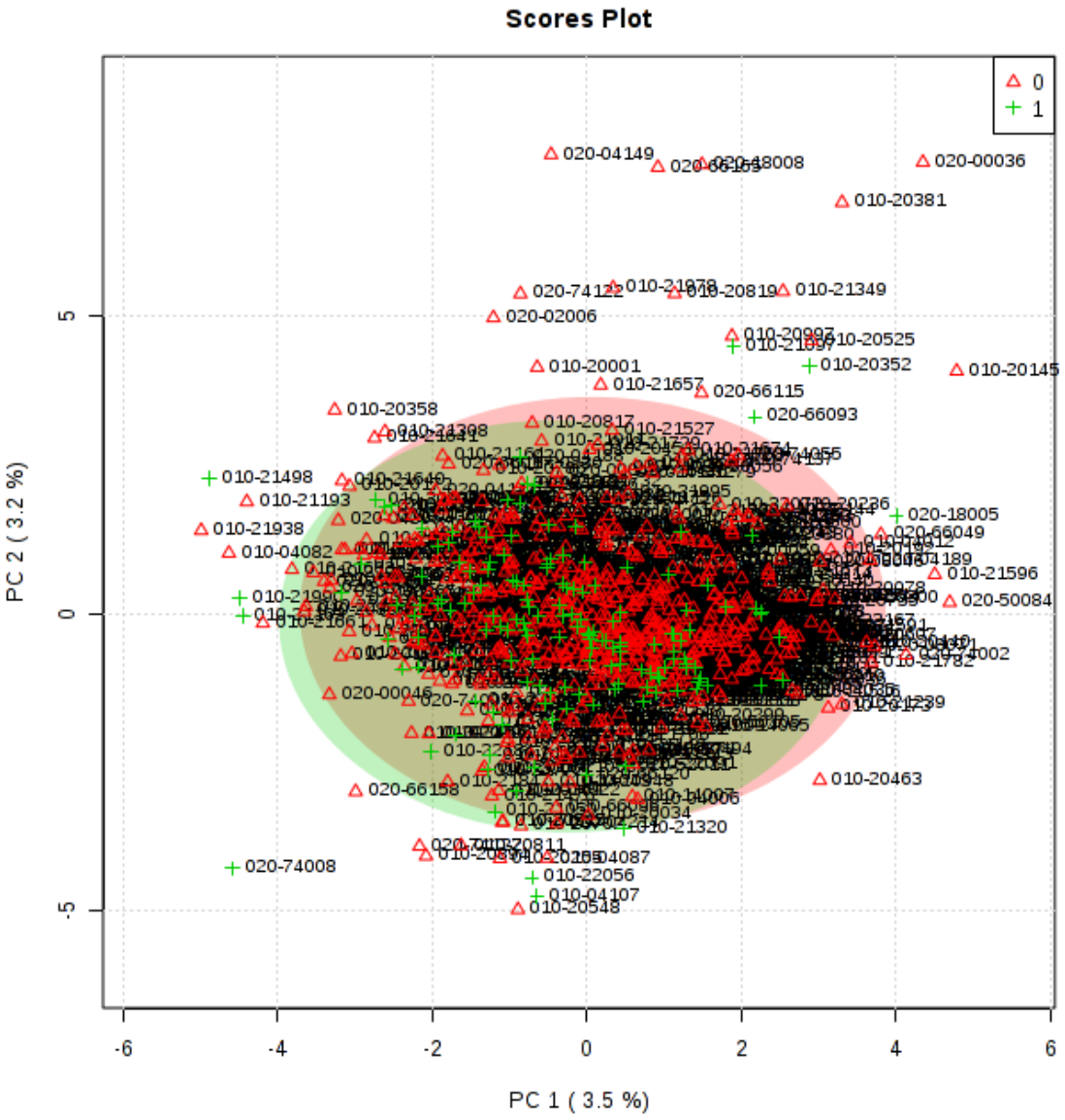


Figure 3.4. Principal component analysis of participants according to their complete dietary data, used to identify potential outliers; 0 – red triangles = controls, 1 – green crosses = GDM cases.

Table 3.3

Participant Characteristics

		Gestational Diabetes Mellitus			
		Yes **N = 160 (17.6 %)**		**No** **N = 749 (82.4 %)**	**P-Value**
BMI (median [lower quartile, upper quartile])		26.2 (24.0, 29.0)		25.0 (22.9, 28.1)	0.003*[a]
Age (mean ± SD)		33 ± 5		30 ± 5	< 0.001*
Ethnicity	Chinese	109 (68.1)	Chinese	406 (54.2)	
N (%)	Malay	18 (11.2)	Malay	216 (28.8)	
	Indian	33 (20.6)	Indian	127 (17.0)	< 0.001*
Education	<Secondary	31 (19.4)	<Secondary	247 (33.0)	
N (%)	Post-Secondary	56 (35.0)	Post-Secondary	277 (37.0)	
	University	73 (45.6)	University	225 (30.0)	< 0.001*
Household monthly income	< 2000	14 (8.8)	< 2000	113 (15.1)	
N (%)	2000 – 6000	92 (57.5)	2000-6000	447 (59.7)	
	> 6000	54 (33.8)	> 6000	189 (25.2)	0.02*
Smoking	No	157 (98.1)	No	728 (97.2)	
N (%)	Yes	3 (1.9)	Yes	21 (2.8)	0.51

		Gestational Diabetes Mellitus			
		Yes		**No**	**P-Value**
		N = 160 (17.6 %)		**N = 749 (82.4 %)**	
Alcohol consumption	No	156 (97.5)	No	715 (95.5)	
N (%)	Yes	4 (2.5)	Yes	34 (4.5)	0.24
Previous history of GDM	No	142 (88.8)	No	737 (98.4)	
N (%)	Yes	18 (11.2)	Yes	12 (1.6)	< 0.001*
Family history of diabetes	No	104 (65.0)	No	533 (71.2)	
N (%)	Yes	56 (35.0)	Yes	216 (28.8)	0.12
Birth order	First child	58 (36.2)	First child	322 (43.0)	
N (%)	Not first child	102 (63.7)	Not first child	427 (57.0)	0.12
Birthweight (g) (mean ± SD)		3067 ± 494		3098 ± 426	0.47
Gestational age at delivery (days) (median [lower quartile, upper quartile])		272 (266, 278)		271 (265, 277)	0.38[a]

Chi-square (2-tailed) tests were conducted to compare group differences, unless otherwise specified.

[a]Mann-Whitney U test was conducted for continuous variables that did not follow a normal distribution

*P < 0.05

3.3.2 Dietary patterns.

Three maternal dietary patterns were identified (Table 3.2): A vegetable-fruit-rice-based diet, high in vegetables, fruit, white rice, bread, low-fat meat and fish, and low in fried potatoes, burgers, carbonated and sugar-sweetened beverages; a seafood-noodle-based diet, high in soup, fish and seafood products, noodles (flavoured and/or in soup), low-fat meat, and seafood, and low in ethnic bread, legumes and pulses, white rice, and curry-based gravies; and a pasta-cheese-processed-meat diet, high in pasta, processed meats, tomato-based and cream-based gravies. These three dietary patterns explained 9.4% of the variation in dietary intake of the participants. Previous studies of dietary patterns in pregnancy have reported between 13.8% and 20.8% variance explained by the first two or three patterns derived by factor analysis or principal component analysis (Brantsaeter et al., 2009; Cucó et al., 2006; Englund-Ögge et al., 2014; Radesky et al., 2007; Thompson et al., 2010; C. Zhang et al., 2006). It was hypothesised from these results that the GUSTO participants had highly variable diets in comparison to the other populations studied.

3.3.3 Dietary patterns and gestational diabetes mellitus.

In an unadjusted logistic regression model, a higher score on the vegetable-fruit-rice-based diet was associated with a higher likelihood of GDM (odds ratio per SD [OR] [95% confidence interval [CI]] = 1.37 [1.16, 1.63], $P < 0.001$). A trend was observed between a higher score on the seafood-noodle-based diet and reduction in the likelihood of GDM, although this did not reach statistical significance (OR [95% CI] = 0.85 [0.73, 1.01], $P = 0.06$; Table 3.4).

After adjustment of the model for confounding variables, the association between GDM and the vegetable-fruit-rice-based diet was no longer statistically significant ($P = 0.33$). However, the association between higher scores on the seafood-noodle-based diet and reduced likelihood of GDM reached statistical significance (OR [95% CI] = 0.77 [0.62, 0.96] $P = 0.02$).

Participants with the highest scores on the seafood-noodle-based diet (highest tertile) were predominantly Chinese, and were less likely to report a family history of diabetes, when compared with women with the lowest compliance to the seafood-noodle-based diet (lowest tertile; Table 3.5).

3.3.4 Validation of dietary patterns.

A statistically significant, moderate correlation was observed between participants' scores on the seafood-noodle-based diet and vegetable-fruit-rice-based diet, with applied scores calculated from the 3-day food diaries (vegetable-fruit-rice-based diet: $R = 0.52$, $P < 0.001$; seafood-noodle-based diet: $R = 0.53$, $P < 0.001$).

Table 3.4

Logistic Regression Analysis of Dietary Patterns with GDM Outcome

	Odds Ratio (OR) (95% CI)	P-value
Vegetable-fruit-rice-based diet		
Unadjusted model	1.37 (1.16, 1.63)	< 0.01*
Multivariate model[1]	1.10 (0.90, 1.35)	0.36
Seafood-noodle-based diet		
Unadjusted model	0.85 (0.73, 1.01)	0.06
Multivariate model[1]	0.74 (0.59, 0.93)	< 0.01*
Pasta-cheese-processed-meat-based diet		
Unadjusted model	0.97 (0.81, 1.16)	0.72
Multivariate model[1]	0.96 (0.79, 1.17)	0.71

[1] Adjusted for pregnancy BMI, age, ethnicity, education, previous GDM, energy intake, and other dietary patterns

*P < 0.05

Table 3.5

Participant Characteristics and Nutrient Composition by Tertile of Seafood-Noodle-Based Diet

		Tertile 1 of Seafood-noodle-based diet (lowest scores)	Tertile 2 of Seafood-noodle-based diet	Tertile 3 of Seafood-noodle-based diet (highest scores)	P-value comparing highest and lowest tertiles
Gestational diabetes mellitus	No	243 (80.2)	255 (84.2)	251 (82.8)	
N (%)	Yes	60 (19.8)	48 (15.8)	52 (17.2)	0.46[a]
Pregnancy BMI (median [lower quartile, upper quartile])		26.1 (23.4, 29.0)	25.3 (23.1, 28.4)	24.7 (22.6, 27.3)	0.095[c]
Age (mean ± SD)		30.2 ± 5.2	30.6 ± 5.2	30.8 ± 5.0	0.11[b]
Ethnicity	Chinese	103 (34.0)	164 (54.1)	248 (81.9)	
N (%)	Malay	84 (27.7)	103 (34.0)	47 (15.5)	
	Indian	116 (38.3)	36 (11.9)	8 (26.4)	< 0.001**[a]
Education	<Secondary	79 (26.1)	104 (34.3)	95 (31.4)	
N (%)	Post-Secondary	98 (32.3)	126 (41.6)	109 (36.0)	
	University	126 (41.6)	73 (24.1)	99 (32.7)	0.05[a]
Household monthly income	< 2000	42 (13.9)	46 (15.2)	39 (12.9)	
N (%)	2000-6000	177 (58.4)	193 (63.7)	169 (55.8)	
	> 6000	84 (27.7)	64 (21.1)	95 (31.4)	0.67[a]

		Tertile 1 of Seafood-noodle-based diet (lowest scores)	Tertile 2 of Seafood-noodle-based diet	Tertile 3 of Seafood-noodle-based diet (highest scores)	P-value comparing highest and lowest tertiles
Smoking	No	292 (96.4)	295 (97.4)	298 (98.3)	
N (%)	Yes	11 (3.6)	8 (2.6)	5 (1.7)	0.20[a]
Alcohol consumption	No	293 (96.7)	288 (95.0)	290 (95.7)	
N (%)	Yes	10 (3.3)	15 (5.0)	13 (4.3)	0.77[a]
Previous history of GDM	No	298 (98.3)	292 (96.4)	289 (95.4)	
N (%)	Yes	5 (1.7)	11 (3.6)	14 (4.6)	0.05[a]
Family history of diabetes	No	198 (65.3)	208 (68.6)	231 (76.2)	
N (%)	Yes	105 (34.7)	95 (31.4)	72 (23.8)	0.003*[a]
Birth order	First child	122 (40.3)	122 (40.3)	136 (44.9)	
N (%)	Not first child	181 (59.7)	181 (59.7)	167 (55.1)	0.25[a]
Birthweight (g) (median [lower quartile, upper quartile])		3045 (2823, 3343)	3100 (2853, 3381)	3120 (2830, 3416)	0.08[c]
Gestational age at delivery (days) (median [lower quartile, upper quartile])		272 (266, 278)	271 (266, 277)	272 (266, 278)	0.17[c]

	Tertile 1 of Seafood-noodle-based diet (lowest scores)	Tertile 2 of Seafood-noodle-based diet	Tertile 3 of Seafood-noodle-based diet (highest scores)	P-value comparing highest and lowest tertiles
Energy (kcal)	1804 (1402, 2177)	1780 (1424, 2193)	1972 (1658, 2356)	0.78[c]
(median [lower quartile, upper quartile])				
<u>Nutrient Composition</u>				
Protein, n% of energy	15.0 (12.5, 17.4)	15.1 (12.8, 17.8)	16.2 (13.5, 18.7)	< 0.001**[c]
(median [lower quartile, upper quartile])				
Total fat, n% of energy	31.5 (25.6, 38.0)	33.0 (27.9, 37.4)	33.0 (29.0, 38.0)	0.01*[c]
(median [lower quartile, upper quartile])				
Carbohydrate, % of energy	52.1 (47.3, 59.6)	51.5 (47.0, 57.1)	50.2 (44.2, 55.6)	< 0.001**[c]
(median [lower quartile, upper quartile])				
Protein (g)	67.8 (49.8, 84.9)	67.0 (50.9, 86.7)	78.0 (64.2, 96.6)	< 0.001**[c]
(median [lower quartile, upper quartile])				

	Tertile 1 of Seafood-noodle-based diet (lowest scores)	Tertile 2 of Seafood-noodle-based diet	Tertile 3 of Seafood-noodle-based diet (highest scores)	P-value comparing highest and lowest tertiles
Carbohydrate (g)	231.7 (184.5, 288.2)	229.5 (179.5, 287.7)	240.1 (195.6, 290.1)	0.12[c]
(median [lower quartile, upper quartile])				
Total fat (g)	63.0 (44.0, 85.2)	64.3 (48.4, 83.2)	70.7 (57.1, 92.9)	< 0.001**[c]
(median [lower quartile, upper quartile])				
<u>Food Group Intake (measured in grams)</u>				
Soup (mean [standard deviation])	16.5 (55.8)	77.6 (119.9)	215.4 (200.6)	< 0.001**[c]
Fish & seafood products (mean [standard deviation])	1.4 (5.7)	6.2 (14.9)	23.4 (37.4)	< 0.001**[c]
Noodles in soup (mean [standard deviation])	4.9 (36.3)	14.7 (45.5)	75.7 (104.9)	< 0.001**[c]
Flavoured noodles (mean [standard deviation])	10.9 (42.2)	42.0 (73.3)	97.3 (126.9)	< 0.001**[c]
Red Meat (B) (mean [standard deviation])	5.7 (17.8)	18 (33.1)	37.1 (48.7)	< 0.001**[c]
Seafood (mean [standard deviation])	2.2 (8.7)	8.8 (20.6)	21.7 (43.8)	< 0.001**[c]
Soya sauce-based gravy (mean [standard deviation])	2.0 (10.1)	3.0 (11.6)	17.0 (53.5)	< 0.001**[c]

	Tertile 1 of Seafood-noodle-based diet (lowest scores)	Tertile 2 of Seafood-noodle-based diet	Tertile 3 of Seafood-noodle-based diet (highest scores)	P-value comparing highest and lowest tertiles
White rice (mean [standard deviation])	195.1 (157.7)	135.9 (142.5)	102.8 (127.6)	< 0.001**[c]
Curry-based gravy (mean [standard deviation])	41.8 (85.2)	10.6 (38.0)	6.1 (32.1)	< 0.001**[c]
Legumes/Pulses (mean [standard deviation])	19.7 (52.6)	3.4 (13.0)	1.3 (7.5)	< 0.001**[c]
Ethnic bread (mean [standard deviation])	38.0 (104.0)	5 (24)	2 (17)	< 0.001**[c]

[a] Chi-square (2-tailed) tests
[b] Independent samples t-test
[c] Mann-Whitney U test
* P < 0.05
** P < 0.001

3.3.5 Nutrient composition between highest and lowest tertile of seafood-noodle-based diet.

Participants in the highest tertile of seafood-noodle-based diet compliance had significantly higher levels of protein and fat consumption (as a percentage of total energy), and lower levels of carbohydrate consumption (Table 3.5).

3.3.6 Sensitivity analysis.

In the subset of participants who had complete information (N = 822), the three dietary patterns derived using varimax rotated factor analysis (Table 3.6; Figure 3.5) closely resembled the patterns produced using the larger dataset (N = 909; Table 3.2). Out of the 10 food groups with the highest factor loadings for each of the three patterns, 80% were the same between the two datasets. A comparison of the results following binary logistic regression in the unimputed dataset with the imputed dataset revealed minor differences in the adjusted models, but a more substantial difference in the unadjusted models (Table 3.7): In the "N=822" dataset, the seafood-noodle-based diet demonstrated a significant association with GDM prior to the adjustment for confounders. This association was not significant in the "N=909" dataset. This implies that the inclusion of N = 87 was affecting the relationship between dietary patterns and GDM prior to imputation. Excluding these women from analysis could risk biasing the results, therefore justifying the inclusion of these individuals in the analysis. Imputation did not appear to significantly change the findings from the adjusted models in the logistic regression analyses.

Table 3.6

Three Dietary Patterns Identified from Exploratory Factor Analysis Using Varimax Rotation (N = 822)

Vegetable-fruit-rice-based diet (3.5% variance explained)		**Seafood-noodle-based diet (3.2% variance explained)**		**Pasta-cheese-processed-meat-based diet (2.7% variance explained)**	
Food Group	*Factor Loading Coefficient*	*Food Group*	*Factor Loading Coefficient*	*Food Group*	*Factor Loading Coefficient*
CLYO vegetables	0.56	Soup	0.46	Pasta	0.78
Other vegetables	0.45	Fish & seafood products	0.41	Tomato-based gravy	0.66
Soup	0.40	Flavoured noodles	0.40	Cheese	0.47
Fruit	0.35	Noodles in soup	0.38	Cream-based gravy	0.46
White rice	0.32	Red Meat (B)	0.33	Meat products	0.26
Red meat (B)	0.30	Seafood	0.28		
Wholemeal/ multigrain bread	0.29	Soya sauce-based gravy	0.26		
Nuts/seeds	0.26	White rice	- 0.32		
Sweetened drinks	- 0.25	Ethnic bread	- 0.44		
Flavoured rice	- 0.29	Legumes/Pulses	- 0.46		
Carbonated drinks	- 0.31				
Burger	- 0.33				
Potatoes (F)	- 0.41				

Extraction Method: Principal Component Analysis
Rotation Method: Varimax with Kaiser Normalization
CLYO vegetables = cruciferous, leafy green, yellow, and orange vegetables
For display purposes, only food groups with a factor loading ≥ 0.25 or ≤ 0.25 are listed in the table.
F = Fried preparation or curry cooked in coconut
B = Boiled, steamed, grilled, roasted, baked, stir fried, braised, or stewed preparation**.**

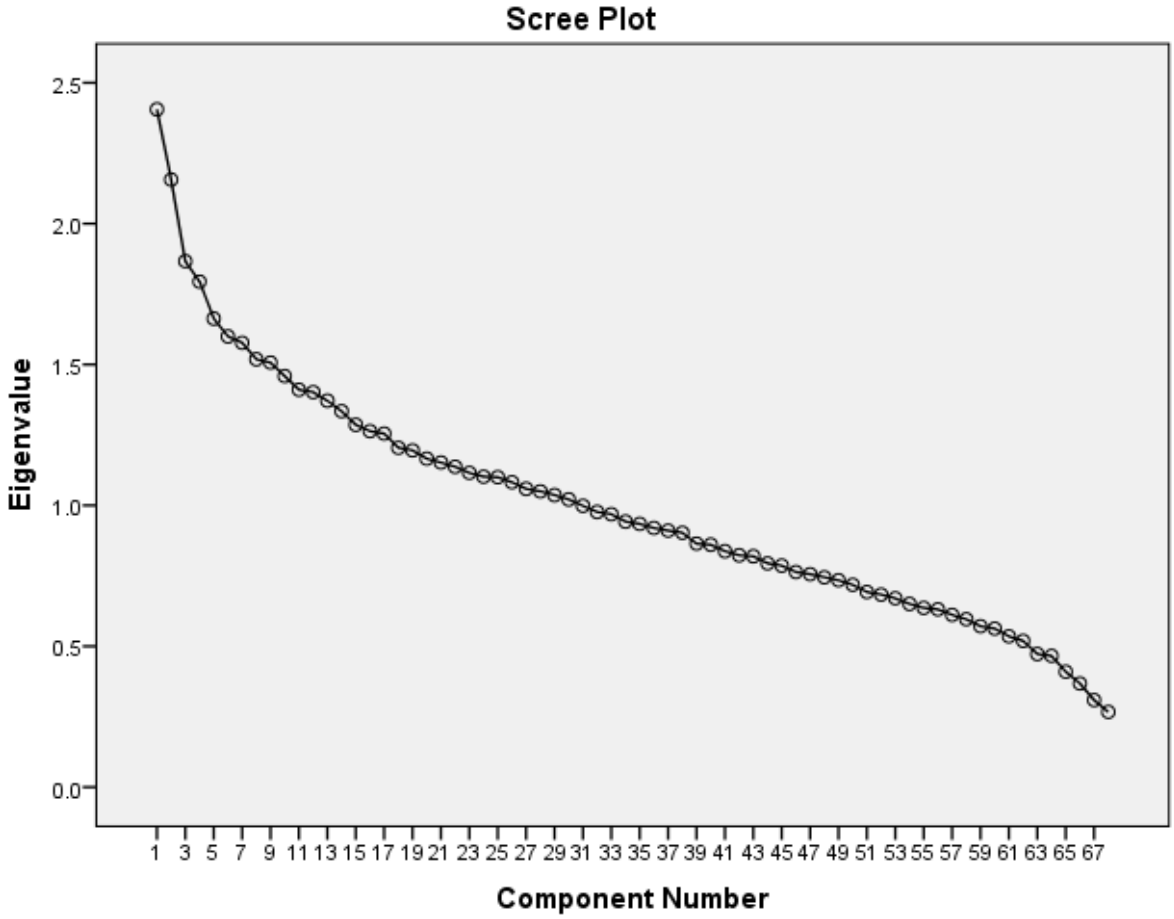


Figure 3.5. Scree plot of eigen values following factor analysis of 68 food groups (N=822)

Table 3.7

Logistic Regression Analysis of Dietary Patterns with GDM Outcome (N = 822)

	Odds Ratio (OR) (95% CI)	P-value
Vegetable-fruit-rice-based diet		
Unadjusted model	1.36 (1.13, 1.62)	< 0.01*
Multivariate model[1]	1.06 (0.86, 1.31)	0.61
Seafood-noodle-based diet		
Unadjusted model	0.80 (0.68, 0.95)	0.01*
Multivariate model[1]	0.75 (0.60, 0.94)	0.01*
Pasta-cheese-processed-meat-based diet		
Unadjusted model	1.01 (0.83, 1.22)	0.96
Multivariate model[1]	0.98 (0.80, 1.21)	0.87

[1] Adjusted for pregnancy BMI, age, ethnicity, education, previous GDM, energy intake, and other dietary patterns

*P < 0.05

3.4 Discussion

This study investigated the association between maternal diet and GDM in a multi-ethnic Asian cohort, using dietary pattern analysis. High scores on the seafood-noodle-based diet were associated with a lower likelihood of GDM, after adjustment for confounding variables. These findings are substantially different to previous studies of maternal dietary patterns related to GDM, predominantly conducted in Western populations.

An explanation for the difference observed between this study and those performed in a Western context may be due to the vast difference in the carbohydrate sources. In Singapore, white rice and noodles (egg and rice) are staple carbohydrates, whereas neither were identified in the maternal dietary patterns constructed using participants from the USA (Radesky et al., 2007; Shin, Lee, & Song, 2015; C. Zhang et al., 2006). White rice and

noodles have different nutritional properties and elicit a different physiological response, particularly in regards to glucose metabolism. Therefore, it is highly plausible that these two major staples in the Singapore diet might be influential in the relationship between maternal diet and GDM in this cohort. In the seafood-noodle-based diet identified, there appears to be an exchange in carbohydrate source across tertiles – with those in the highest tertile consuming more noodles and less white rice. Previous research has alluded to a detrimental effect of rice intake in Asian populations, linking higher consumption with an increased risk of type 2 diabetes (Nanri et al., 2010; Villegas et al., 2007). White rice has a high glycaemic index (Sugiyama, Tang, Wakaki, & Koyama, 2003), particularly when compared with noodles. A high-glycaemic-index diet during pregnancy has been associated with GDM (Hodge, English, O'Dea, & Giles, 2004; Mohan et al., 2009; Salmerón et al., 1997).

The only other study reported to have investigated dietary patterns and GDM in an Asian context was conducted in China by He et al. (2015). They found that those participants who were in the highest tertile of their "vegetable" dietary pattern consumption had a reduced likelihood of GDM, whereas those in the highest tertile of their "sweets and seafood" dietary pattern had an increased likelihood of GDM. The interesting difference between the findings of He et al. (2015) and this study conducted on a Singapore cohort was that the vegetable-rich dietary pattern derived using the GUSTO cohort was not associated with GDM. Additionally, in the Singaporean cohort, seafood was a major component of the dietary pattern found to be associated with a reduced likelihood of GDM. This may be explained by the difference in the foods consumed in combination. In the Singapore cohort, the seafood was accompanied by soup and noodles, whereas in the Chinese cohort the seafood was accompanied by Cantonese desserts and sugar-sweetened beverages. This difference reflects the uniqueness of food consumption in similar ethnic groups in different countries, and highlights the need for caution when generalising findings outside of the setting they are discovered in.

Dietary assessment using 24-hr recalls has some limitations. Dietary recalls are subject to recall bias, and social desirability bias (particularly when administered face-to-face). Social desirability bias is apparent in almost all self-reported dietary assessment and is difficult to control for. Women who reported dietary intake with energy values > 2 standard deviations below the mean (< 800kcal) were deemed to be underreporting and were excluded from analyses. In an attempt to ameliorate recall bias, 24-hr recalls were interviewer-administered. Although 24-hr recalls have been shown to be an effective form of dietary assessment in a population setting, they do not capture the full picture of the dietary intake of an individual,

and the food choice can be mutually exclusive in a 24-hr period (dietary variety in some cases may not be assessed to completeness within just 24 hr). For this reason, a sub-set of three-day food diaries were collected from participants in the week following OGTT (N = 208) to validate the two dietary patterns of interest that were generated from the 24-hr recalls. A statistically significant, moderate correlation was observed between participants' scores on the seafood-noodle-based diet ($R = 0.52$, $P < 0.001$) and vegetable-fruit-rice-based diet ($R = 0.53$, $P < 0.001$), with applied scores calculated from the three-day food diaries. By this criterion, the 24-hr dietary recalls adequately represented typical dietary pattern consumption of the participants in this study. Previous work on the GUSTO dietary data has also demonstrated the reproducibility of 24-hr recalls and therefore justifies the use of dietary pattern construction using this dietary assessment method (Chong et al., 2015). Future work in this area should consider the use of validated food frequency questionnaires for dietary assessment, as they would be more likely to represent a longer length of dietary exposure.

The main strength of this study was the ability to investigate dietary pattern associations with GDM using a large multi-ethnic Asian cohort. Using a cohort with close to 1000 participants greatly increased the ability to detect differences and to adequately represent the population of interest – Singaporean expectant mothers. The GUSTO cohort was also used to address the gap in the literature surrounding maternal dietary patterns and GDM in Asian populations, where very little research has been conducted. In addition, the dietary pattern approach performed in this study to investigate the relationship between diet and GDM allowed identification of food patterns that are typically consumed in the population, and to consider the synergistic impact of nutrients consumed in combination, in their typical food matrices.

3.5 Conclusion

To summarise, the current study investigated the association between maternal dietary patterns and GDM in a multi-ethnic Asian cohort. The study findings demonstrated that in this cohort, compliance to a seafood-noodle-based diet was associated with a reduced likelihood of GDM. This finding is novel, and has not been reported previously. The difference observed in these findings as compared to previous findings, predominantly in Western settings, demonstrates that very little is known or understood about the relationship between diet and GDM in this at-risk group. It is clear that further research should be conducted in Asian populations to better understand the role of maternal diet in GDM development and progression. This research domain would benefit not only from

observational studies to validate the findings of this study, but also mechanistic studies that can elucidate the metabolic mechanisms linking associations between the maternal dietary patterns and GDM.

Chapter Four: The plasma Metabolome in GDM

4.1 Introduction

Diagnosis of GDM typically occurs late in the second trimester using an OGTT (American Diabetes Association, 2011). An OGTT consists of the pregnant woman ingesting a high glucose load (usually 50 g, 75 g, or 100 g) dissolved in solution, to test her carbohydrate response. However, this test is highly unpleasant for many mothers, resulting in nausea and discomfort following the glucose load, and it is also conducted late in pregnancy (usually between the 24th and 26th weeks of gestation). Diagnosing GDM later in pregnancy risks exposing the fetus to an aberrant environment resulting from a mother's high blood glucose levels, established prior to clinical diagnosis. The best way to overcome this risk is to prevent GDM from developing. Before prevention strategies can be employed, however, women who are at risk for developing GDM first need to be identified. To date, no tool to predict GDM in early pregnancy has been effectively employed in routine clinical practice. Correa Vargas, Sen, & Illanes (2014) concluded that prediction tools used thus far have lacked efficacy, and that further research is needed to identify a tool for accurate prediction of GDM. A study conducted in Singapore also found that high-risk screening (presence of: obesity [BMI of > 30.0 kg/m^2], family history of type 2 diabetes [first-degree relative], GDM in a previous pregnancy, previous delivery of a baby with a birth weight ≥ 4.5 kg, and/or ethnic origin with known high prevalence of diabetes) failed to detect 50% of women who developed GDM, deeming it inadequate for prediction (Chong et al., 2014). One approach that has been employed in the search for biomarkers of GDM is metabolomics.

Metabolomics attempts to characterise the metabolome – the comprehensive set of low molecular weight compounds (<1500Da) present in a biological system (Villas-Bôas, 2007). Metabolites include endogenously derived compounds such as amino acids, fatty acids, nucleic acids, hormones, enzymes, amines, pigments, vitamins, organic acids, short peptides, and sugars, and can also include compounds from exogenous sources such as environmental contaminants, food additives, toxins, and drug-related compounds (Wishart et al., 2013). The metabolome is the smallest biological unit downstream of genomics, transcriptomics, and proteomics, and is influenced by diet, lifestyle, environment, and physical activity, thus making the metabolome the closest representation of the phenotype of an individual. This "snapshot" of phenotype is valuable in the search for biomarkers of GDM. In addition, the metabolome provides information that could allow for the elucidation of metabolic

mechanisms underlying GDM development, which would assist in steering the direction of research into prevention strategies, such as dietary interventions.

Despite preliminary findings (Table 1.6), metabolomics studies in GDM have failed to consistently identify a set of metabolites clinically useful for the prediction of GDM development. Studies to date have predominantly been conducted on small, heterogeneous sample populations, which are more prone to random effects. Exploring the metabolome in early pregnancy offers the opportunity to discover biomarkers of disease development. However, because metabolomic studies of GDM are still in the early stages, and findings are inconsistent, it would be beneficial to first identify the metabolome associated with established disease, to better understand the metabolic dysfunction directly related to GDM development. Understanding the effect of GDM on the metabolome will help to ensure that biomarkers identified in early pregnancy have biological relevance and to help understand their trajectory to established disease. Large, well-conducted cohort studies are required to confirm the metabolomic profile associated with GDM development.

The following study investigated the maternal plasma metabolome, with the aim of identifying a metabolomic signature of GDM, using participants from a large multi-ethnic Asian pregnancy cohort.

4.2 Methods

4.2.1 Participants.

The plasma metabolome was characterised for a sub-set of participants from the GUSTO cohort. The details of the cohort and associated ethics approval for this work have been described in Chapter 2.

GUSTO participants were eligible for this study of the plasma metabolome if they met the following criteria:

a) Underwent an OGTT at 26-28 weeks' gestation;
b) Had more than 200 µl of plasma available in the −80º C biobank;
c) Had complete data for maternal age, BMI, and ethnicity (as these are three major clinical predictors of GDM).

4.2.2 Plasma collection.

Maternal plasma is a suitable candidate for metabolomic investigations of GDM because metabolite levels in the blood represent a combination of many processes in the body, and are therefore thought to reflect an overall snapshot of the state of metabolic functioning (Gowda et al., 2008). Plasma is also less affected by external influences, such as non-nutrient metabolites resulting from cooking processes and recent dietary intake, than urine (Gibney et al., 2005). As this study was interested in both metabolic mechanisms and potential biomarkers, plasma remained the preferred choice of biological specimen. Plasma was collected from participants at the time of the OGTT, following a strict standard operating procedure (SOP; Appendix 3). The samples analysed in this study were collected after an overnight fast. Analysing the metabolome from plasma collected after an overnight fast reduced the likelihood of identifying differences based on recent dietary intake, which can complicate the interpretation of results related to disease processes (Yin, Lehmann, & Xu, 2015). Blood was collected in 10-ml collection tubes containing the anti-coagulant ethylenediaminetetraacetic acid (EDTA), with the addition of trasylol. Concerns have been raised about the effect of EDTA on the metabolome, particularly in LC-MS analysis (Gonzalez-Covarrubias, Dane, Hankemeier, & Vreeken, 2013). However, in GC-MS, EDTA does not have a substantial effect on the metabolome, and has a much smaller effect than the most common alternative used in plasma collection tubes, heparin (Bando et al., 2010). Trasylol is a brand of apotinin, an enzyme inhibitor (Zyznar, 1981), which, when added to blood collection tubes, can reduce the likelihood of protein degradation prior to freezing the samples. The addition of trasylol, in combination with strict operating procedures, ensures that the sample analysed closely reflects the state of metabolism in the participant at the time of blood collection. Collection tubes were kept on wet ice prior to the separation of plasma from whole blood. Processing did not exceed 4 hr in total. Yin et al. (2013) have previously demonstrated that the EDTA plasma metabolome does not change significantly after 4 hr on ice, but they did observe significant effects if stored at room temperature. The blood was centrifuged at 3000 RPM for 10 min at 4 ºC to separate the plasma out from whole blood. The supernatant (plasma) was then transferred into 400 µl aliquots. The aliquots were sealed and frozen at −80 ºC until required. Plasma samples were shipped on dry ice from Singapore to New Zealand in January 2015, and were stored at −80 ºC until analysis. The samples were randomised according to a stratified randomisation sequence that ensured cases of GDM were spread equally throughout the analytical batches, so that technical effects would not confound

the results. The randomisation order was constructed by a biostatistician who was not directly involved in the analysis of the samples, to reduce the possibility of bias. Researchers directly involved in the analysis remained blind to the GDM status of the participants throughout the experiment and data pre-processing.

4.2.3 Metabolomic profiling of plasma.

Methods for metabolomic profiling of plasma were adapted from those published in *Nature Protocols* by Dunn, Broadhurst, Atherton, Goodacre, & Griffin (2011) and Smart, Aggio, Van Houtte, & Villas-Bôas (2010).

4.2.3.1 Internal standards.

Eight internal standards (Sigma-Aldrich) were used: alanine-d_4 (99%, 10 mM), tryptophan-d_5 (99%, 10 mM), hexanoic acid-d_{11} (99%, 10 mM), phenylalanine-d_5 (99%, 10 mM), tyrosine-d_2 (99%, 10 mM), octanoic acid-d_{15} (99%, 10 mM), stearic acid-d_{35} (99%, 10 mM), and citric acid-d_4 (99%, 10 mM). Alanine-d_4, tyrosine- d_2, phenylalanine- d_5, hexanoic acid- d_{11}, citric acid- d_4, and tryptophan- d_5 were dissolved in 36 ml of milli-Q water (internal standard mix 1). Octanoic acid-d_{15} and stearic acid- d_{35} were dissolved in 36 ml of 100% methanol (LC-MS grade, Merck; internal standard mix 2). The intended purpose of multiple internal standards was to adjust for variation introduced throughout the experimental process – and, most importantly, for any loss of sample. For this work, multiple internal standards were chosen over a single internal standard, as a single internal standard does not behave the same as all metabolites likely to be identified in the samples, particularly if they are not in the same class of compound. The eight internal standards chosen were selected based on their retention time, ensuring a spread of internal standards over the course of the GC-MS run.

Internal standard mix 1 (20 µl) and internal standard mix 2 (20 µL) were added to pre-chilled 1.5 ml Eppendorf microcentrifuge tubes. After plasma samples were defrosted at 4 ºC, 200 µl from each sample was aliquoted into the pre-chilled tubes containing the two internal standard mixes. The mixture was vortexed for 30 s. A 40 µl aliquot of the sample-IS mix was then transferred from each tube into a pre-chilled 50 ml falcon tube which remained on dry ice throughout the experiment, to form a pooled quality control (QC) mix. The QC mix was homogenised and used for monitoring reproducibility throughout the analysis.

The sample for each participant was evaporated to dryness using a SpeedVac Concentrator (Savant SC250EXP, ThermoFisher Scientific) with a −104 °C refrigerated vapour trap (Savant RVT4104, ThermoFisher Scientific) at 0.8 hPa with no heating, until it was dry

(drying times for each sample are listed in Appendix 4). Dried samples were stored at −80 ºC. Once all samples were dried, the pooled QC mix was defrosted at 4 ºC, vortexed, and measured into 200 µl aliquots that were added to microcentrifuge tubes containing internal standards. The aliquots were evaporated to dryness as per the samples, and then placed into the −80 ºC storage until extraction.

4.2.3.2 Extraction.

Metabolites were extracted from plasma prior to GC-MS analysis. Extraction was performed for three purposes: enzyme inactivation, precipitation of high molecular weight compounds, and extraction of low molecular weight compounds. Cold methanol extraction was selected as the extraction approach as it has been used successfully to extract metabolites from serum and plasma, and has been found to be a simple and effective means for protein precipitation; Want et al. (2005) compared 14 different extraction methods for serum and found that methanol extraction resulted in the lowest amount of residual protein, and was found to be highly reproducible (Want et al., 2005). No single extraction protocol is capable of extracting the full range of metabolites from a biological specimen, and, as such, the use of a single extraction method limits the range of metabolites extracted from the samples. Methanol, being a polar solvent, is capable of solubilising polar metabolites, and therefore methanol extraction limits metabolites extracted to predominantly polar species such as purines and pyrimidines, nucleotides, organic acids, hydrophilic vitamins, and amino acids. However, methanol extraction has also been shown to extract some nonpolar metabolites such as fatty acids and phospholipids (Dunn et al., 2011). The extraction procedure employed in this study was based on the published protocol by Dunn et al. (2011), with some minor modifications. The amount of plasma available for use in the current study was less than what was analysed in Dunn et al. (2011). Despite this difference, the methanol to sample ratio was retained. To maximise the number of metabolites extracted using a single extraction solvent, rather than using one concentration of methanol (as per Dunn et al., 2011), the current study used three different ratios of methanol to milli-Q water, sequentially decreasing in polarity. Using three solutions with different polarities allowed the extraction of a larger range of metabolites. Another small modification to the protocol was that rather than using the set speed-vac drying time reported, the current study dried the samples to completeness and recorded the drying time.

Dried samples and QCs were taken from the −80 ºC freezer and placed on wet ice. Three solutions of methanol and milli-Q water (50%, 80%, and 100% MeOH to milli-Q) were

prepared and stored at −20 °C. Cold MeOH: H2O (50%, 500 µL) was added to the samples and they were vortexed for 10 s and left on wet ice for 30 min to aid resuspension. The samples were vortexed for a further minute, then centrifuged at 3500 RPM for 5 min at −4 °C (Eppendorf 5430R, Mediray, New Zealand). The supernatant from each sample was transferred to a pre-chilled 1.5 ml Eppendorf microcentrifuge tube. The pellet was resuspended in 500 µl of cold 80% MeOH:H2O solution and the same procedure was followed, omitting the 30-min stand-time. The supernatant was combined with the supernatant from the 50% methanol extraction. Finally, the same extraction process was repeated for 100% methanol. The combined supernatants were then evaporated to dryness using a SpeedVac Concentrator (Savant SC250EXP, ThermoFisher Scientific) with a −105 °C refrigerated vapour trap (Savant RVT4104, Thermo Scientific) at 0.8 hPa, with no heating (drying times for each sample are listed in Appendix 4). Once dry, samples were stored at −80 °C.

4.2.3.3 Derivatisation.

Derivatisation of metabolites is required when performing GC-MS analysis as the metabolites need to be adequately volatile and thermally stable to be analysed by gas chromatography (Dettmer, Aronov, & Hammock, 2007). The two main types of derivatisation employed are alkylation and silylation. Methyl chloroformate (MCF) was chosen for derivatisation of metabolites in this study – an alkylation reaction that converts amino and other organic acids into volatile carbamates and esters (Villas-Bôas, Smart, Sivakumaran, & Lane, 2011). Silylation is effective for analysing alcohols, sugars and their derivatives, and basic carboxylic acids but has been found to be less reproducible than alkylation for analysing amino acids and other organic acids, which are important biological components when investigating metabolic diseases such as GDM (Villas-Bôas et al., 2011). In addition to the types of compounds MCF derivatisation can target, and the improved reproducibility of MCF derivatisation when compared with N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) derivatisation (silylation reaction), alkylation also helps to preserve the quality of the GC column when using biological fluids, when compared to silylation. This is because the silylation reaction leaves a large amount of non-derivatized and non-volatile compounds as well as derivatizing reagents in the sample which can also be injected into the column (Villas-Bôas et al., 2011). These factors accumulate and can compromise the quality of the results. The derivatisation procedure used in this study was based on that reported in Smart et al. (2010), with modifications. The modifications included a prolonged rest period following

the initial resuspension of samples and QCs with NaOH, as the time reported in Smart et al. was insufficient to adequately resuspend the dried pellet. The MCF reaction reagent volumes from Smart et al. were doubled in this study as human plasma has a high concentration of metabolites and as such requires the addition of more MCF to ensure there is enough available to derivatise all metabolites present. The double MCF method has been used previously for derivatisation of human serum for GC-MS analysis (de Seymour et al., 2014).

Dried samples and QCs were resuspended with 400 µL of 1M NaOH, vortexed for 20 s and rested on ice for 20 min before being vortexed for 10 s. The resuspended samples were each transferred to a silanised glass test tube. Methanol (334 µL) and pyridine (68 µL) were added, followed by vortexing for 10 s. For the alkylation reaction, the tube containing sample was placed in the vortexer and remained on it throughout the reaction. Methyl chloroformate (40 µL MCF, Merck, Germany) was added to the sample and the mixture was vortexed vigorously for 30 s. A second 40µL of MCF was added, and vortexing continued for 30 s. With the sample still on the vortex, 400 µL of chloroform was added and vortexing continued for a further 10 s. Finally, 800 µL of 50 mM $NaHCO_3$ was added to terminate the reaction, and the mixture was vortexed for a final 10 s. Derivatised samples were centrifuged for 3 min at 2000 RPM to separate the aqueous and organic phases (Eppendorf 5430R, Mediray, New Zealand). The upper aqueous phase was removed and residual water in the organic phase was removed by addition of 300 mg of anhydrous Na_2SO_4. The organic layer was then transferred from the silanised tube to a glass vial with insert, for GC-MS analysis.

4.2.3.4 Gas chromatography-mass spectrometry (GC-MS) analysis.

Each sample was analysed using GC-MS. The GC-MS instrument parameters were derived from Smart et al., with modifications to the signal threshold, inlet liner, inlet mode, and max isothermal time in order to increase sensitivity, improve reproducibility, and prevent carryover. The GC-MS used for analysis in this study was an Agilent 7890B gas chromatograph coupled to a 5977A inert mass spectrometer. One microlitre of sample was injected using an Agilent autosampler into a glass split/splitless 4 mm ID single taper inlet liner packed with deactivated glass wool (Restek). The inlet was set to 290 °C, splitless at 180 kPa for 1 min, column head pressure 57 kPa, with a constant column flow of 1.0 mL/min, giving a calculated average initial linear velocity of 35 cm/s. Purge flow was set to 25 mL/min, 1.1 min after injection. The column was a fused silica ZB-1701 30 m long, 0.25 mm i.d., 0.15 µm stationary phase (86% dimethylpolysiloxane, 14% cyanopropylphenyl, Phenomenex) with a 5 m guard column. Carrier gas was instrument grade helium (99.9999%,

BOC). GC oven temperature programming started isothermally at 45 °C for 2 min, increased 9 °C/min to 180 °C, held 5 min, increased 40 °C/min to 220 °C, held 5 min, increased 40 °C/min to 240 °C, held 11.5 min, increased 40 °C/min to 280 °C, and held 10 min in order to prevent carryover of late-eluting compounds such as cholesterol. Total run time was 48 min. The transfer line to the MSD was maintained at 250 °C, the source at 250 °C and quadrupole at 130 °C. The detector was turned on 5.5 min into the run to avoid detector overload from eluting solvent. The detector was run in positive-ion, electron-impact ionisation mode, at 70 eV electron energy, with electron multiplier set with no additional voltage relative to the autotune value. Identification of compounds was carried out using mass spectra acquired in scan mode from 38 to 550 amu, with detection threshold of 50 ion counts. The low threshold was used to increase sensitivity and thus reduce the number of missing values in the data matrix.

Each GC-MS run took 48 min, excluding cool-down time, therefore, < 30 injections could be performed every 24 hr. This limitation determined the size of the analytical batch, i.e., the number of samples derivatised and run on the GC-MS each day. There was a total of 27 injections/ 24 hr. The injection order remained consistent each working day (Figure 4.1), with the exception of Mondays, and after septum and liner changes. On Mondays, when the GC-MS had not been running consistently over the weekend (and also when any part of the instrument was cleaned or replaced), an additional two injections of the first QC were carried out, to better condition the column before plasma samples were injected.

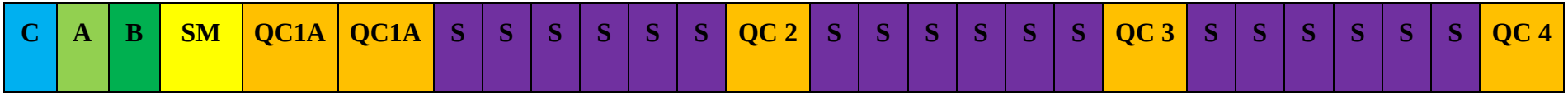


Figure 4.1. Shows the daily injection sequence.

QC = Quality Control samples were used to track variation in plasma metabolites as a result of batch variation over the duration of the GC-MS analysis. A = Alkane mix (C7-C30 saturated alkane mix purchased from Sigma Aldrich, St. Louis, USA) was used to track the variation of the GC-MS instrument over time. B = Blank/analytical negative control which was prepared from the beginning of the experiment and treated the same as a sample but without the addition of plasma. This was used to measure contamination from solvents, equipment, and human error throughout the processing steps. C = Chloroform was used to clear any build-up of contamination on the column that originated from the septum, helium supply, and inlet liner. STM = Standard Mix which consisted of L - alanine (1 mM), L - leucine (1 mM), L - lysine (monohydrochloride) (1 mM), aspartic acid (1 mM), L - tryptophan (1 mM), L - cysteine (1 mM), 2 - hydroxybutyric acid (1 mM), citric acid (1 mM), and 3 - methyl - 2 - oxovaleric acid (sodium salt) (0.86 mM). The SM underwent MCF derivatisation with the samples and was used to track any variation due to derivatisation over batches. S = Plasma Sample (randomised and blinded).

Visual inspections of the quality control sample chromatograms (QCs, alkane mix, chloroform, blank, and standard mix) were performed at the end of each day and samples were rerun if there were any discrepancies. Septum and liner replacement occurred after every 200 injections, unless required earlier.

4.2.4 Data processing.

Raw GC-MS files underwent two phases of data extraction to derive maximal metabolite detection from the data. Phase 1 used the in-house library for compound identification: Each raw GC-MS Agilent format data file was converted to a common data format (CDF) file for compatibility with cross-platform software for data extraction. The CDF files were run through an automated deconvolution and identification system (AMDIS) - a freely available, downloadable software (Stein, 1999). AMDIS was used in conjunction with a University of Auckland in-house metabolite library developed by Silas Villas-Bôas, which contains retention times and mass spectra for 217 reference compounds (listed in Appendix 5.). AMDIS settings were determined by optimising different deconvolution and identification parameters on 10 QC samples selected from regular intervals throughout the duration of the GC-MS runs (parameters chosen are listed in Appendix 6). Settings were chosen that would optimise the number of peaks identified without compromising the identification accuracy. These settings were then applied to all of the samples, QCs, and controls for batch deconvolution and identification. An AMDIS report was generated. AMDIS generates a high number of missing values, so a further step was carried out to improve the sensitivity of data extraction. MassOmics, version 2.3, a windows-based data extraction application developed by Ting-Li Han for in-house R scripts and the XCMS package (Smith, Want, O'Maille, Abagyan, & Siuzdak, 2006), was used to generate a composite list of all metabolites detected in the dataset. This report contained metabolite identifications, mass spectral identification scores, the most abundant ion for each library match, the number of times each metabolite was detected in the whole data set, and the amount of retention time drift for each metabolite. The summary report was curated manually in Excel to remove incorrect peak-picking, and reference ions that lay outside the acquisition window. XCMS was used to extract the peak height of the discriminatory ion for each retention time bin for each metabolite, whether or not AMDIS had detected it in the original analysis. This enabled a significant reduction in the number of zero values, an increase in sensitivity at low abundance, with only a modest increase in false positive identification.

Phase 2 consisted of determining putative identifications using the National Institute of Standards & Technology 2014 (NIST14) mass spectral library. Due to the limited number of metabolites in the in-house library, not all peaks in the chromatograms were identified in Phase 1. To address this limitation, the NIST14 mass spectral library was used to increase the number of peak identifications and therefore metabolites for downstream analysis. The NIST14 library contained mass spectra from 242,466 metabolites (V. Shen, Siderius, Krekelberg, & Hatch, 2017). As AMDIS is a small 32-bit programme, it cannot work with the whole NIST14 library in batch mode, so a data extraction method developed by Dr Elizabeth McKenzie was used. A subset of the NIST14 library was created using NIST14 library match output from Agilent Chemstation software for QC samples representatively selected from the GC-MS runs. The Chemstation integration parameters were set to be sensitive to very low abundance compounds, and search-match parameters were expansive, with 10-15 compound options for each feature included, in order to produce a generous subset mass spectral library of 6000 compounds. Following the development of the subset library, the procedure used for Phase 1 was repeated, but with slightly different settings for AMDIS (Appendix 6), replacing the in-house library with the sub-set NIST14 library. After the completion of Phases 1 and 2, the two datasets were merged into a single Excel spreadsheet.

4.2.5 Data cleaning.

The dataset underwent checking and correction before normalisation. Data cleaning consisted of the following steps:

1) Removal of any metabolites identified in less than 50 injections (samples and QCs combined), as this was 50% of the smallest experimental group in this study.
2) Removal of contaminants that were defined as either a) less than 80% of sample peak heights for that metabolite above the peak height in the blanks AND b) fold change between samples and blanks less than 5. Metabolites that met either a) OR b) were graphed in samples vs blanks for visual inspection to determine whether they were contaminants.
3) Resolution of duplicate identifications – in some instances the same peak was picked more than once, as more than one metabolite. To determine which peak was more likely, the frequency in the dataset and library match factor of the metabolite were assessed in a probabilistic fashion and the peaks inspected visually if its identity was not resolved using these parameters.

4) Visual inspection of overlaid chromatograms – a stratified random selection of 50 samples and 50 QCs were used to generate overlaid chromatographic peaks for every metabolite. Any overlays that appeared to be capturing multiple peaks in the same retention time window underwent visual checking of the chromatogram and re-integration using a corrected retention time window. Re-integration was performed using the ion extractor package in the MassOmics software, developed by Raphael Aggio (Aggio, Villas-Bôas, & Ruggiero, 2011).

4.2.6 Statistical analysis.

4.2.6.1 Normalisation of data and identification of outliers.

Normalising metabolomic data is a difficult task due to the sheer diversity of detected compounds. Eight internal standards were used to normalise samples for variation introduced as part of the experimental process. To determine which internal standard to use for normalisation of each metabolite, Pearson's product-moment correlation was performed on the QC data for each of the metabolites with each of the eight internal standards. The internal standard with the highest correlation to each metabolite was selected for normalisation of that metabolite in the samples. To normalise by internal standard the following equation was performed:

$$\frac{\text{Metabolite peak height in a sample}}{\text{internal standard (highest correlated with that metabolite) in the sample}} \times \text{median of the chosen internal standard in the QCs}$$

Any metabolites that did not have a correlation with an internal standard higher than a Pearson's $R > 0.9$ did not undergo internal standard normalisation. Correlations were used as a cut-off because a low correlation indicated that the internal standard was not behaving in the same way as the metabolite of interest throughout the course of the study. If a metabolite was to be normalised using an internal standard that was not behaving in the same way, then this could introduce error into the dataset rather than removing variation caused as a result of the experimental process.

To correct for residual technical variation, a batch correction normalisation strategy was applied. Two strategies were tested. The strategy that passed a Kruskal Wallis test to prove that the strategy removed batch effects (no longer significant [$P > 0.05$] across the different days of GC-MS analysis [Days 1- 61]) was used to normalise the data.

The two strategies tested were:

a) median centering of the samples across batches using the pooled QCs
b) median centering of the samples across batches using the samples themselves

To perform median centering the following process was implemented:

Step 1) calculate the centering factor (either using QCs or samples):

$$\frac{\text{median value of metabolite X in a batch}}{\text{median of that metabolite across all samples}}$$

Step 2) repeat this calculation for each of the 61 batches for metabolite X;

Step 3) repeat this process for each metabolite;

Step 4) apply the centering factor specific to the metabolite and corresponding batch, to the samples;

Step 5) multiply the newly generated values for each metabolite by the median of that metabolite in all samples (or QCs if using the QC approach) – this brings the normalised values back to a number scaled to the peak height that was initially generated.

4.2.6.2 Missing values and outliers.

The final data contained 0.02% missing values. Missing values were imputed as half of the minimum value detected in the samples (Xia, Psychogios, Young, & Wishart, 2009). The metabolite scores were standardised to Z-scores and then a principal-components analysis (PCA) was performed to identify outliers. A visual inspection of the PCA plot containing all participants, conducted using Metaboanalyst version 3.0, was used to distinguish outliers. Outliers were removed so that they did not heavily influence the downstream analyses.

4.2.6.3 Analysis of metabolite differences between groups.

Two approaches were used to test for significant differences between GDM cases and non-GDM controls in the cohort. All analyses were performed using SPSS v23.0. P-value < 0.05 was considered as statistically significant. The data was tested for normality using the Shapiro-Wilk test in SPSS v23.0 (Shapiro & Wilk, 1965).

1) Non-parametric Mann-Whitney U test for comparison of medians between GDM cases and non-GDM controls was performed, as the metabolites did not follow a normal distribution and normality was not reached after transformation attempts.

2) Binary logistic regression was performed for each metabolite to determine if there was a significant relationship between the metabolite and GDM status, in an adjusted model which accounted for confounding variables. Confounding variables were determined as demographic variables that were significantly different between GDM cases and non-GDM controls: age, ethnicity, education (also as a proxy for household monthly income as they were highly correlated [Pearson's R = 0.573 P-value < 0.001]), skinfold measurements (also a proxy for BMI as they were highly correlated [Pearson's R = 0.773 P-value < 0.001]), and previous GDM.

Receiver operator characteristic (ROC) curves were plotted using Metaboanalyst version 3.0 (http://www.metaboanalyst.ca/) to illustrate the ability of the metabolites identified to correctly differentiate GDM cases from controls. ROC curves were generated by Monte-Carlo cross validation (MCCV) using balanced subsampling. In each MCCV, two thirds (2/3) of the samples were used to evaluate feature importance. Classification models were validated on the 1/3 samples that were left out. The procedure was repeated multiple times to calculate the performance and confidence interval of each model (Metaboanalyst, 2016). The classification method used was linear support vector machine (SVM). Area under the curve (AUC) was calculated and reported for each ROC curve produced. The AUC results from the metabolite ROC curves were compared with ROC curves produced using clinical risk factors currently applied in practice to identify women at risk of GDM.

4.3 Results

4.3.1 Participants.

A total of 793 women met the inclusion criteria listed in Section 4.2. After removal of outliers according to the principal component analysis (Figure 4.2), a total of 787 participants remained in the study for further analysis. The characteristics of the participants in this study are listed in Table 4.1. Missing values for discrete demographic variables were imputed by a multiple imputation procedure (Rubin, 1996) using SPSS for birth order (N = 10), education level (N = 8), monthly household income (N = 49), family history of diabetes (N = 28), previous history of GDM (N = 19), smoking (N = 1), and alcohol consumption (N = 20). Participants were from three ethnic groups, with both of their parents belonging to the same ethnic group: Chinese (57.2%), Malay (25.4%), and Indian (17.4%). The average age of the participants was 31 (± 5) years and the median (lower quartile, upper quartile) BMI was 25.3 (23.1, 28.2). Of 787 participants, 17.8% (N = 140) were diagnosed with GDM using WHO

criteria (Alberti & Zimmet, 1998); fasting plasma glucose >7 mmol/L or 2 h post-load plasma glucose ≥7.8 mmol/L). Women diagnosed with GDM were more likely to have a higher BMI and body fat (skinfold measures), advanced age, history of GDM in a previous pregnancy, attained a higher education, and receive a higher household income level.

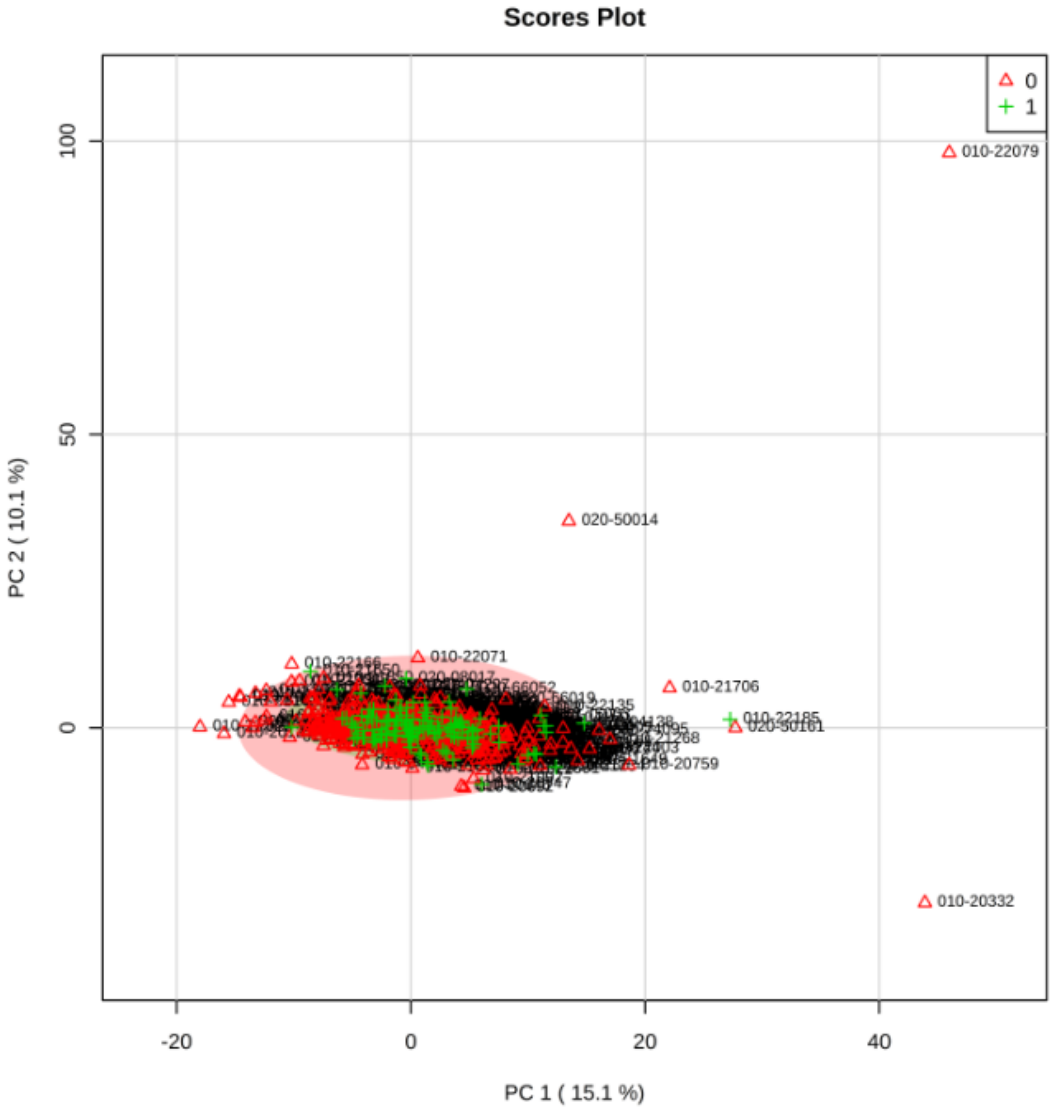


Figure 4.2. Principal component analysis of participants according to their complete metabolite profile revealed six outliers; 0 – red triangles = controls, 1 – green crosses = GDM cases.

Table 4.1

Participant Characteristics

		Gestational Diabetes Mellitus			
		No		Yes	P-Value
		N= 647 (82.2 %)		N=140 (17.8 %)	
BMI (median [lower quartile, upper quartile])		25.0 (22.9, 27.9)		26.3 (24.2, 29.2)	0.02*
Sum of skinfolds (mean ± SD)		106.4 (21.9)		112.8 (22.1)	0.002*
Age (mean ± SD)		30 ± 5		33 ± 5	< 0.001*
Ethnicity	Chinese	356 (55)	Chinese	94 (67.1)	
N (%)	Malay	185 (28.6)	Malay	15 (10.7)	
	Indian	106 (16.4)	Indian	31 (22.1)	< 0.001*
Education	<Secondary	213 (33.3)	<Secondary	27 (19.3)	
N (%)	Post-Secondary	229 (35.8)	Post-Secondary	43 (30.7)	
	University	197 (30.8)	University	70 (55.0)	< 0.001*
Household monthly income(SGD)	< 2000	99 (16.1)	<2000	13 (9.9)	
N (%)	2000-6000	339 (55.8)	2000-6000	67 (51.5)	
	> 6000	170 (28.1)	>6000	51 (38.9)	0.02*
Smoking	No	629 (97.2)	No	136 (97.8)	
N (%)	Yes	18 (2.8)	Yes	3 (2.2)	0.99

		Gestational Diabetes Mellitus			
		No		Yes	P-Value
		N= 647 (82.2 %)		N=140 (17.8 %)	
Alcohol consumption	No	616 (97.8)	No	136 (99.3)	
N (%)	Yes	14 (2.2)	Yes	1 (0.7)	0.49
Previous history of GDM	No	618 (98.3)	No	123 (88.5)	
n (%)	Yes	11 (1.7)	Yes	16 (11.5)	< 0.001*
Family history of diabetes	No	438 (70.4)	No	89 (64.5)	
N (%)	Yes	184 (29.6)	Yes	49 (35.5)	0.19
Birth order	First child	277 (43.4)	First Child	52 (37.1)	
N (%)	Not first child	361 (56.6)	Not first child	88 (62.9)	0.19
Birthweight (g) (mean ± SD)		3100 ± 423		3087 ± 489	0.76
Gestational age at delivery (days) (median [lower quartile, upper quartile])		272 (266, 278)		271 (265, 277)	0.07

Chi-square (2-tailed) tests were conducted to compare group differences of categorical variables, Independent t-test was conducted for continuous variables that followed a normal distribution, and Mann-Whitney U test was conducted for continuous variables that did not follow a normal distribution *P < 0.05

4.3.2 Metabolites identified.

A total of 219 compounds were identified in the plasma samples. The list of identified compounds, the library they were derived from, and their match factor with the library standard can be found in Appendix 7. Of the 219 compounds, 34.2% were identified (80-100% match), 31.1% were putatively identified (60-79% match), and 34.7% were unknown (>60% match). The most common classes of compounds identified or putatively identified were lipid species or fatty acids (30%), amino acids or derivatives (20%), and carboxylic acids (17%).

4.3.3 Significant differences in metabolites between GDM and non-GDM groups.

A Mann-Whitney U test to compare differences in median metabolite levels between GDM cases and non-GDM controls resulted in 59 metabolites of significance, after the adjustment for multiple comparisons using a Benjamini-Hochberg technique (Benjamini & Hochberg, 1995); $P < 0.05$ and False discovery rate [FDR] < 0.05); Table 4.2). However, after adjusting for confounders (age, skinfolds, ethnicity, education, and previous GDM), 47 metabolites were significantly associated with GDM ($P < 0.05$), and 26 metabolites remained significant after adjustment for multiple comparisons (FDR < 0.05; Table 4.3 & Figure 4.3).

Table 4.2

Metabolites Detected as Significantly Different Between Cases and Controls After Adjustment for Multiple Comparison Testing

Metabolite	P-Value	False Discovery Rate (FDR)	Fold Change (healthy/GDM)	Coefficient of variation
2-Hydroxybutyric acid	4.33E-15	9.48E-13	0.35	0.14
Glutamic acid	2.36E-08	2.19E-06	0.52	0.13
Methyl L-beta-hydroxyisobutyrate	3E-08	2.19E-06	0.51	0.16
Glutaric acid, decyl 4-(trifluoromethyl) benzyl ester	4.74E-08	2.59E-06	0.55	0.14
2-Oxovaleric acid	9.44E-08	4.13E-06	0.38	0.28
Citric acid	2.18E-07	7.94E-06	0.54	0.07
Glutaric acid	3.12E-07	9.77E-06	0.48	0.27
1,1-Dimethyl-2-[2-(methoxycarbonyl) prop-2-yl] hydrazine	6.41E-07	1.76E-05	0.27	0.15
3-Methyl-2-oxopentanoic acid	1.55E-06	3.76E-05	0.62	0.13
4-Methyl-2-oxopentanoic acid	2.86E-06	6.27E-05	0.59	0.12
1-Propene-1,2,3-tricarboxylic acid, trimethyl ester	5.45E-06	1.09E-04	0.62	0.13
Cis-aconitic acid	8.25E-06	1.51E-04	0.47	0.08
Heptadecanoic acid	1.65E-05	2.77E-04	0.72	0.07
l-Norvaline, N-(2-	3.05E-05	4.50E-04	0.54	0.17

Metabolite	P-Value	False Discovery Rate (FDR)	Fold Change (healthy/GDM)	Coefficient of variation
methoxyethoxycarbonyl)-, propyl ester				
l-Leucine, N-methoxycarbonyl-, ethyl ester	3.08E-05	4.50E-04	0.59	0.20
Palmitoyl chloride	3.33E-05	4.56E-04	0.58	0.20
4-Pyrimidinamine, 6-methyl-	4.76E-05	6.08E-04	0.57	0.16
l-Valine, N-methoxycarbonyl-, pentyl ester	5.19E-05	6.08E-04	0.54	0.18
1-Oxaspiro[2.5]octane, 5,5-dimethyl-4-(3-methyl-1,3-butadienyl)-	5.73E-05	6.08E-04	0.66	0.09
Arachidonic acid	5.77E-05	6.08E-04	0.63	0.11
7,11-Hexadecadienal	5.83E-05	6.08E-04	0.63	0.11
o-Tolunitrile, alpha-cyano-	6.31E-05	6.12E-04	0.57	0.16
Ethyl 8-amino-6-methoxycinchoninate	6.42E-05	6.12E-04	0.57	0.17
Lactic acid	6.77E-05	6.17E-04	0.60	0.09
1H-1,3-Benzimidazol-4-amine, 5-methyl-	8.65E-05	7.37E-04	0.58	0.17
1,2,3,4-Tetrahydrobenzimidazo[2,1-b] quinazolin-12-ol	8.79E-05	7.37E-04	0.57	0.16

Metabolite	P-Value	False Discovery Rate (FDR)	Fold Change (healthy/GDM)	Coefficient of variation
Hexadecanoic acid, 14-methyl-, methyl ester	9.08E-05	7.37E-04	0.64	0.16
l-Alanine, N-methoxycarbonyl-, ethyl ester	1.10E-04	0.001	0.57	0.19
Unknown 045.100.59.89.0.74.67.8.	1.61E-04	0.001	0.61	0.10
Guanidine	1.63E-04	0.001	0.62	0.16
Acetamide, 2-anilino-N,N-dimethyl-	2.26E-04	0.002	0.61	0.17
Hexanoic acid	3.25E-04	0.002	0.64	0.55
1,2-Benzenediol, o-(4-methoxybezoyl)-o'-(5-chlorovaleryl)-	4.13E-04	0.003	0.59	0.20
Ethyl 6-methyl-3-oxopentadecanoate	0.001	0.004	0.74	0.14
2H-naphth[2,1-e]-1,3-oxazine-2,4(3H)-dione	0.001	0.004	0.66	0.14
D-Alanine, N-ethoxycarbonyl-, heptyl ester	0.001	0.004	0.59	0.15
Pyruvic acid	0.001	0.004	0.77	0.90
11,14-Eicosadienoic acid	0.001	0.004	0.74	0.14
Methylenedioxyamphetamine acetate	0.001	0.005	0.67	0.12

Metabolite	P-Value	False Discovery Rate (FDR)	Fold Change (healthy/GDM)	Coefficient of variation
Itaconic acid	0.001	0.005	0.62	0.18
Cyclobutene, 2-propenylidene-	0.001	0.007	1.83	0.22
Creatinine	0.002	0.009	0.74	0.18
Indan, 1-methyl-	0.002	0.01	0.72	0.23
1,2,3-Trimethyl-cyclopent-2-enecarboxaldehyde	0.002	0.01	0.72	0.78
Hydrazine, (3-methoxyphenyl)-	0.002	0.01	0.60	0.52
9-Heptadecenoic acid	0.003	0.01	0.76	0.13
1-(2,3-Dihydro-indol-1-yl)-2-(9-methyl-9H-1,3,4,9-tetraaza-fluoren-2-ylsulfanyl)-ethanone	0.003	0.02	0.63	0.22
Unknown 128.100.42.10.5.129.6.8.	0.004	0.02	0.64	0.17
Citric acid secondary peak	0.004	0.02	0.78	0.75
(7H) Thiopyrano [3,4-c] isoxazole, 3,3a,4,5-tetrahydro-7-(1-methylethyl)-	0.004	0.02	0.61	0.21
Eicosapentaenoic acid (EPA)	0.004	0.02	0.60	0.21
Pentadecanoic acid	0.005	0.02	0.75	0.12
Malic acid	0.005	0.02	0.60	0.15
1-Aminocyclopentanecarboxylic acid, N-methoxycarbonyl-, octyl	0.005	0.02	0.55	0.13

Metabolite	P-Value	False Discovery Rate (FDR)	Fold Change (healthy/GDM)	Coefficient of variation
ester				
Docosahexaenoic acid	0.007	0.03	0.85	0.18
(E)-Tetradec-2-enal	0.009	0.03	0.78	0.07
Glycine, N-ethyl-N-(2-methoxyethoxycarbonyl)-, 2-methoxyethyl ester	0.01	0.04	0.70	0.11
Hexadecanoic acid, 15-methyl-, methyl ester	0.01	0.04	0.83	0.17
Decane	0.01	0.04	0.89	0.14
Glutathione	0.01	0.05	0.63	0.63
Phenylalanine	0.01	0.05	0.90	0.09
Stearic acid	0.01	0.05	0.76	0.09
Benzene, 2-heptenyl-, (Z)-	0.02	0.05	0.87	0.19

Table 4.3

Metabolites Detected as Significantly Associated with GDM After Adjustment for Confounding Variables

Metabolite	CAS number	Match factor (%)	Odds Ratio	95% CI	P-Value	FDR	Direction of association
2-Hydroxybutyric acid	5094246	99	1.89	(1.53, 2.34)	<0.001	7.04E-07	Higher in GDM
Citric acid	77929	96	1.60	(1.30, 1.98)	<0.001	0.001	Higher in GDM
Glutaric acid, decyl 4-(trifluoromethyl) benzyl ester		58	1.55	(1.26, 1.91)	<0.001	0.002	Higher in GDM
1-Propene-1,2,3-tricarboxylic acid, trimethyl ester	20820773	75	1.52	(1.23, 1.87)	<0.001	0.006	Higher in GDM
Methyl L-beta-hydroxyisobutyrate	80657574	89	1.46	(1.19, 1.79)	<0.001	0.01	Higher in GDM
Ethyl 8-amino-6-methoxycinchoninate	858466656	40	1.36	(1.14, 1.61)	0.001	0.02	Higher in GDM
3-Methyl-2-oxopentanoic acid	3715319	94	1.41	(1.15, 1.72)	0.001	0.02	Higher in GDM
Lactic acid	10326417	99	1.35	(1.13, 1.62)	0.001	0.03	Higher in GDM
Malic acid peak 2	6915157	79	1.35	(1.12, 1.63)	0.002	0.03	Higher in GDM

Metabolite	CAS number	Match factor (%)	Odds Ratio	95% CI	P-Value	FDR	Direction of association
l-Alanine, N-methoxycarbonyl-, ethyl ester	88406400	78	1.34	(1.12, 1.60)	0.002	0.03	Higher in GDM
1,2,3,4-Tetrahydrobenzimidazo [2,1-b] quinazolin-12-ol	24415632	41	1.33	(1.11, 1.59)	0.002	0.03	Higher in GDM
2-Oxovaleric acid	1821029	81	1.32	(1.11, 1.58)	0.002	0.03	Higher in GDM
Palmitoyl chloride	112674	63	1.35	(1.12, 1.62)	0.002	0.03	Higher in GDM
l-Norvaline, N-(2-methoxyethoxycarbonyl)-, propyl ester		72	1.37	(1.12, 1.67)	0.002	0.03	Higher in GDM
1H-1,3-Benzimidazol-4-amine, 5-methyl-		38	1.33	(1.11, 1.59)	0.002	0.03	Higher in GDM
Malic acid	6915157	70	1.36	(1.11, 1.66)	0.003	0.04	Higher in GDM
Unknown 128.100.42.10.5.129.6.8.	-	69	1.32	(1.10, 1.59)	0.003	0.04	Higher in GDM
1-(2,3-Dihydro-indol-1-yl)-2-(9-methyl-9H-1,3,4,9-tetraaza-fluoren-2-ylsulfanyl)-ethanone	352328940	53	1.30	(1.09, 1.56)	0.004	0.04	Higher in GDM

Metabolite	CAS number	Match factor (%)	Odds Ratio	95% CI	P-Value	FDR	Direction of association
9-Octadecenamide	301020	64	0.32	(0.15, 0.69)	0.004	0.04	Lower in GDM
Heptadecanoic acid	506127	56	1.36	(1.11, 1.67)	0.004	0.04	Higher in GDM
Hydrazine, (3-methoxyphenyl)-	15384391	51	1.33	(1.10, 1.60)	0.004	0.04	Higher in GDM
4-Methyl-2-oxopentanoic acid	4502005	96	1.33	(1.11, 1.62)	0.004	0.04	Higher in GDM
Guanidine	80706	59	1.30	(1.09, 1.55)	0.004	0.04	Higher in GDM
Itaconic acid	97654	74	1.33	(1.10, 1.61)	0.004	0.04	Higher in GDM
1,2-Benzenediol, o-(4-methoxybezoyl)-o'-(5-chlorovaleryl)-		52	1.30	(1.08, 1.55)	0.005	0.04	Higher in GDM
l-Valine, N-methoxycarbonyl-, pentyl ester	1046143003	86	1.30	(1.08, 1.56)	0.005	0.04	Higher in GDM
Acetamide, 2-anilino-N,N-dimethyl-	14307892	66	1.28	(1.07, 1.52)	0.007	0.05	Higher in GDM
l-Leucine, N-methoxycarbonyl-, ethyl ester	88406433	90	1.28	(1.07, 1.54)	0.007	0.06	Higher in GDM
7,11-Hexadecadienal	849947713	69	1.32	(1.07, 1.62)	0.008	0.06	Higher in GDM

Metabolite	CAS number	Match factor (%)	Odds Ratio	95% CI	P-Value	FDR	Direction of association
Arachidonic acid	506321	-	1.31	(1.07, 1.61)	0.009	0.07	Higher in GDM
1-Aminocyclopentanecarboxylic acid, N-methoxycarbonyl-, octyl ester	626040	80	1.26	(1.05, 1.51)	0.01	0.09	Higher in GDM
Glycine, N-ethyl-N-(2-methoxyethoxycarbonyl)-, 2-methoxyethyl ester		82	1.38	(1.07, 1.79)	0.02	0.10	Higher in GDM
Succinic acid, 2,2,3,3-tetrafluoropropyl 3-methylbut-2-yl ester		88	1.26	(1.04, 1.52)	0.02	0.10	Higher in GDM
(E)-Tetradec-2-enal	51534362	75	1.29	(1.05, 1.58)	0.02	0.10	Higher in GDM
Pyruvic acid	127173	86	1.24	(1.03, 1.49)	0.02	0.13	Higher in GDM
Unknown 045.100..59.89.0...74.67.8.	-	93	1.32	(1.04, 1.67)	0.02	0.13	Higher in GDM
Adipic acid	124049	69	1.26	(1.03, 1.54)	0.02	0.13	Higher in GDM
2H-naphth[2,1-e]-1,3-oxazine-2,4(3H)-dione		51	1.23	(1.02, 1.49)	0.03	0.16	Higher in GDM

Metabolite	CAS number	Match factor (%)	Odds Ratio	95% CI	P-Value	FDR	Direction of association
Glutaric acid, decyl 4-(trifluoromethyl)benzyl ester		58	1.21	(1.02, 1.44)	0.03	0.16	Higher in GDM
Methylenedioxyamphetamine acetate	36209719	57	1.26	(1.02, 1.54)	0.03	0.16	Higher in GDM
Glycine, N-(methoxyoxoacetyl)-, methyl ester	89464631	57	1.21	(1.02, 1.45)	0.03	0.17	Higher in GDM
(7H)Thiopyrano[3,4-c]isoxazole, 3,3a,4,5-tetrahydro-7-(1-methylethyl)- (peak 1)	128869512	47	1.22	(1.02, 1.46)	0.03	0.18	Higher in GDM
Alanine	56417	95	1.24	(1.02, 1.52)	0.04	0.18	Higher in GDM
Creatinine	60275	93	1.24	(1.01, 1.52)	0.04	0.21	Higher in GDM
Glutathione	70188	87	1.23	(1.01, 1.49)	0.04	0.21	Higher in GDM
Cyclobutene, 2-propenylidene-	52097855	33	0.76	(0.58, 1.00)	0.05	0.22	Higher in GDM
Quinazoline, 4-(1-methylethyl)-, 1-oxide	50915327	55	1.28	(1.00, 1.64)	0.05	0.22	Higher in GDM

Confounding variables adjusted in logistic regression model: Age, skinfolds, ethnicity, education, previous GDM

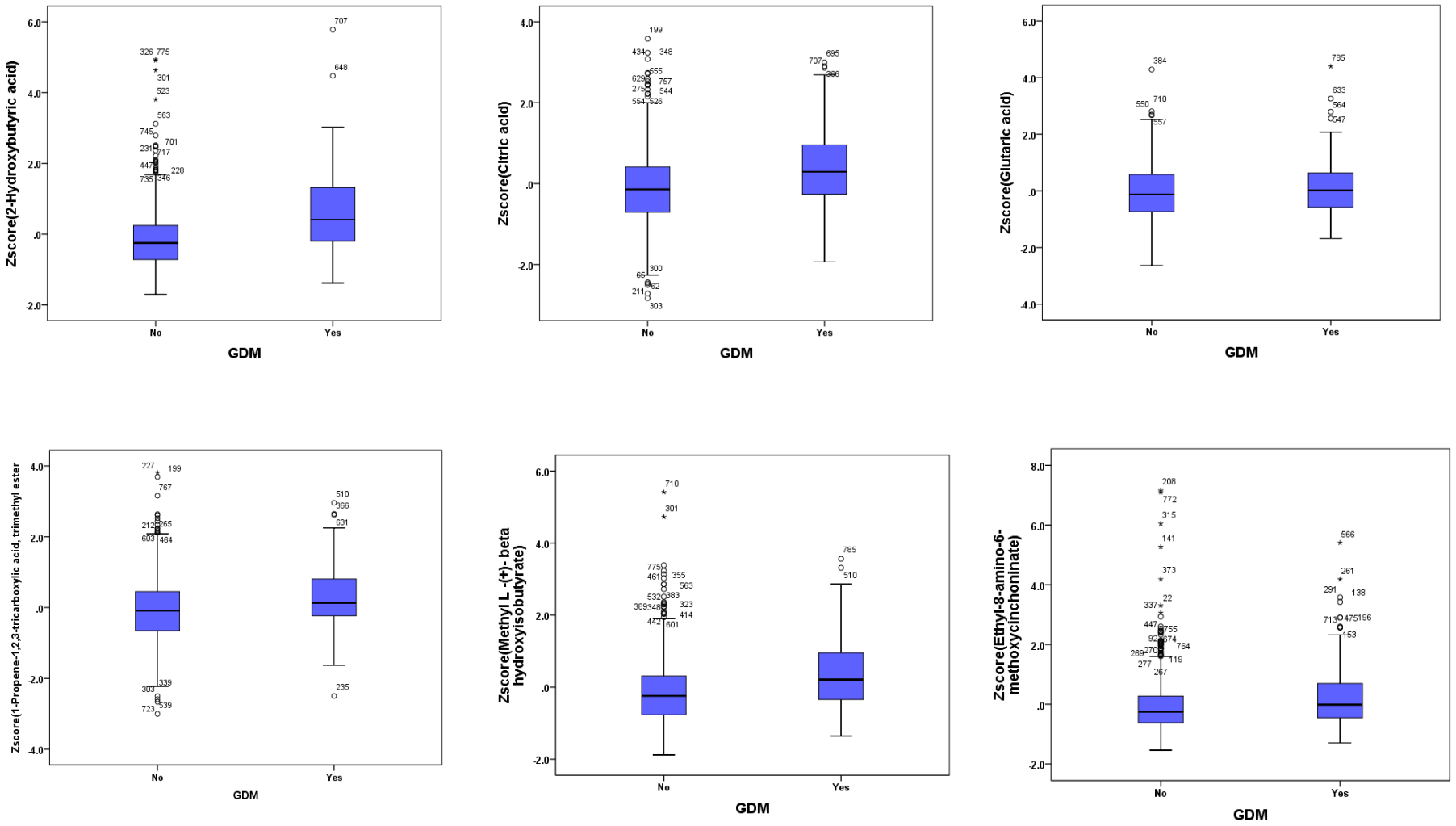


Figure 4.3. Boxplots of the 26 metabolites that remained significantly different between GDM and non-GDM after adjustments for confounding variables and multiple comparison testing.

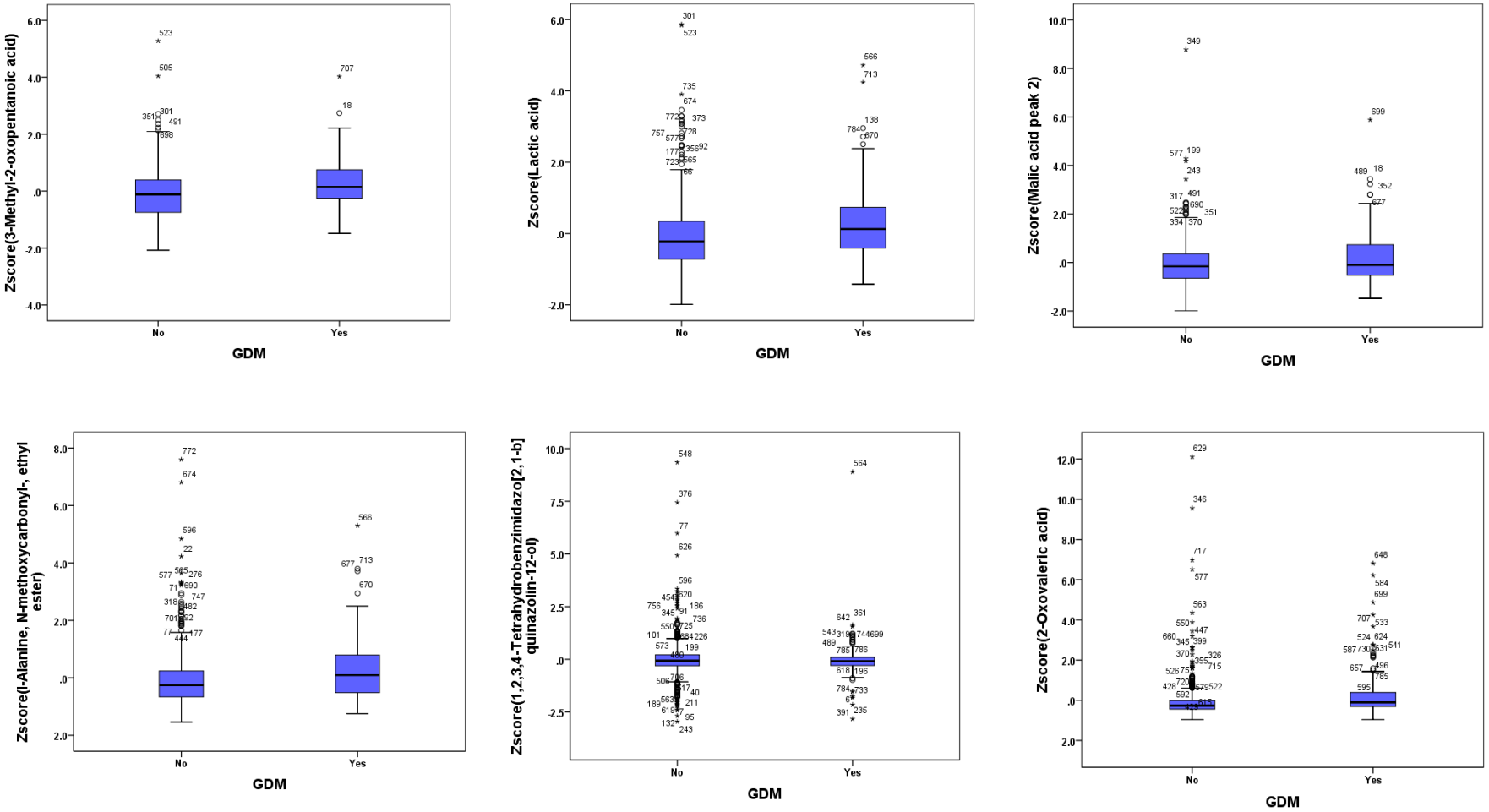


Figure 4.3 continued… Boxplots of the 26 metabolites that remained significantly different between GDM and non-GDM after adjustments for confounding variables and multiple comparison testing.

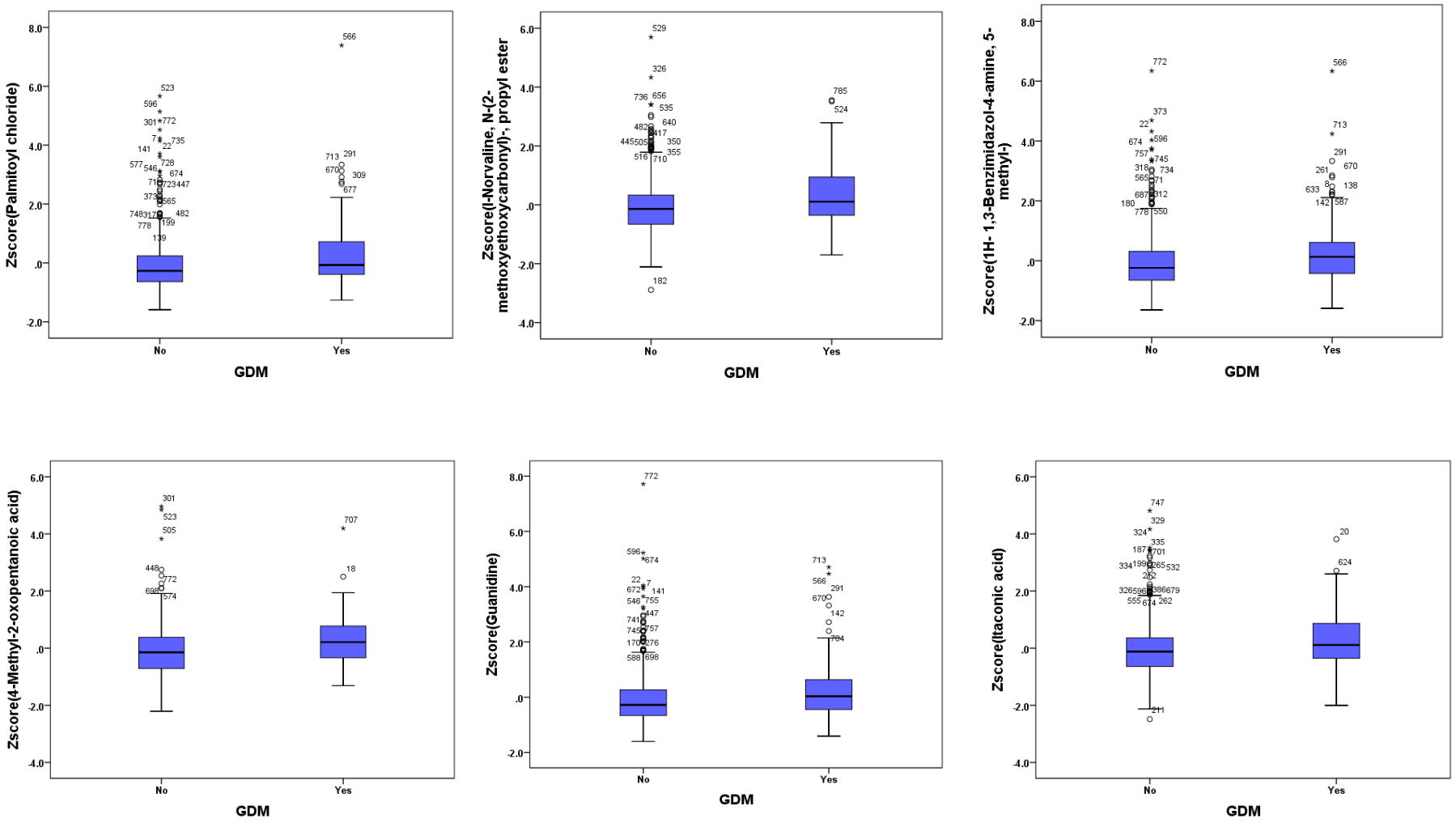


Figure 4.3 continued… Boxplots of the 26 metabolites that remained significantly different between GDM and non-GDM after adjustments for confounding variables and multiple comparison testing.

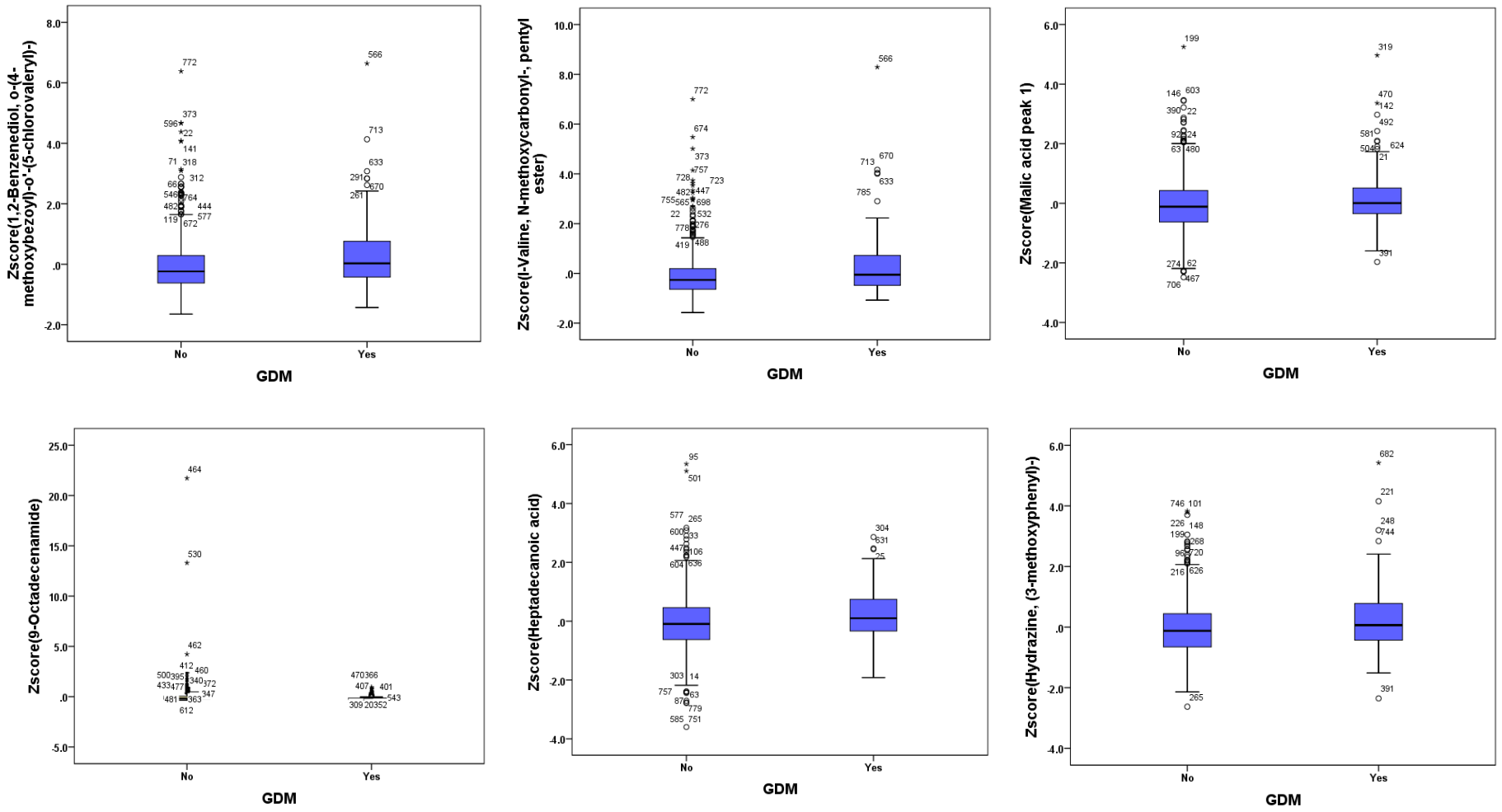


Figure 4.3 continued… Boxplots of the 26 metabolites that remained significantly different between GDM and non-GDM after adjustments for confounding variables and multiple comparison testing.

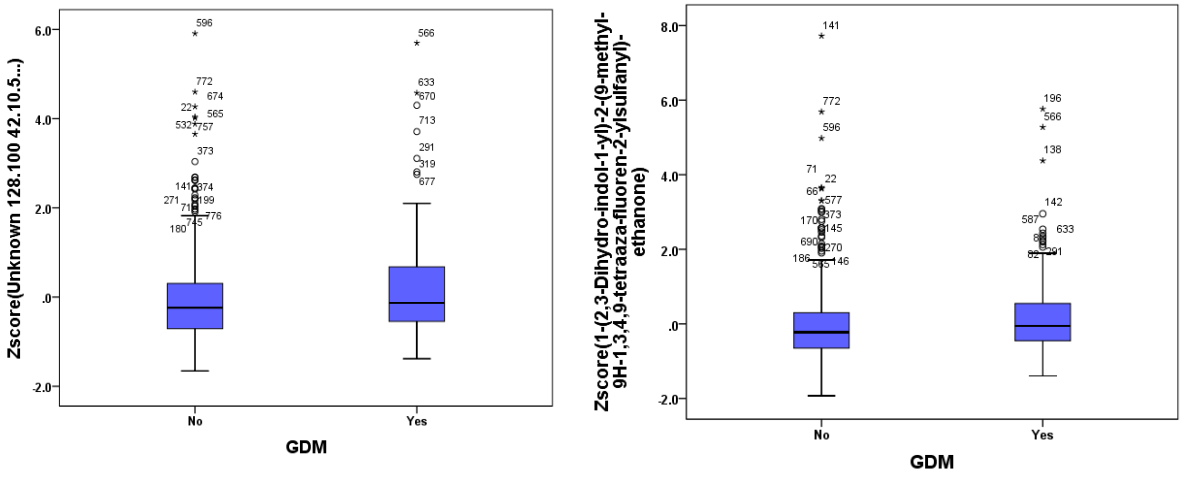


Figure 4.3 continued… Boxplots of the 26 metabolites that remained significantly different between GDM and non-GDM after adjustments for confounding variables and multiple comparison testing.

4.3.4 Receiver operating characteristic curves.

The metabolite that was most significantly different between cases and controls after adjustment for confounding variables, 2-hydroxybutyric acid, was entered into a ROC curve to determine how well it could correctly differentiate GDM classification. The resulting AUC was 0.71 (95% CI: 0.69, 0.76; Figure 4.4).

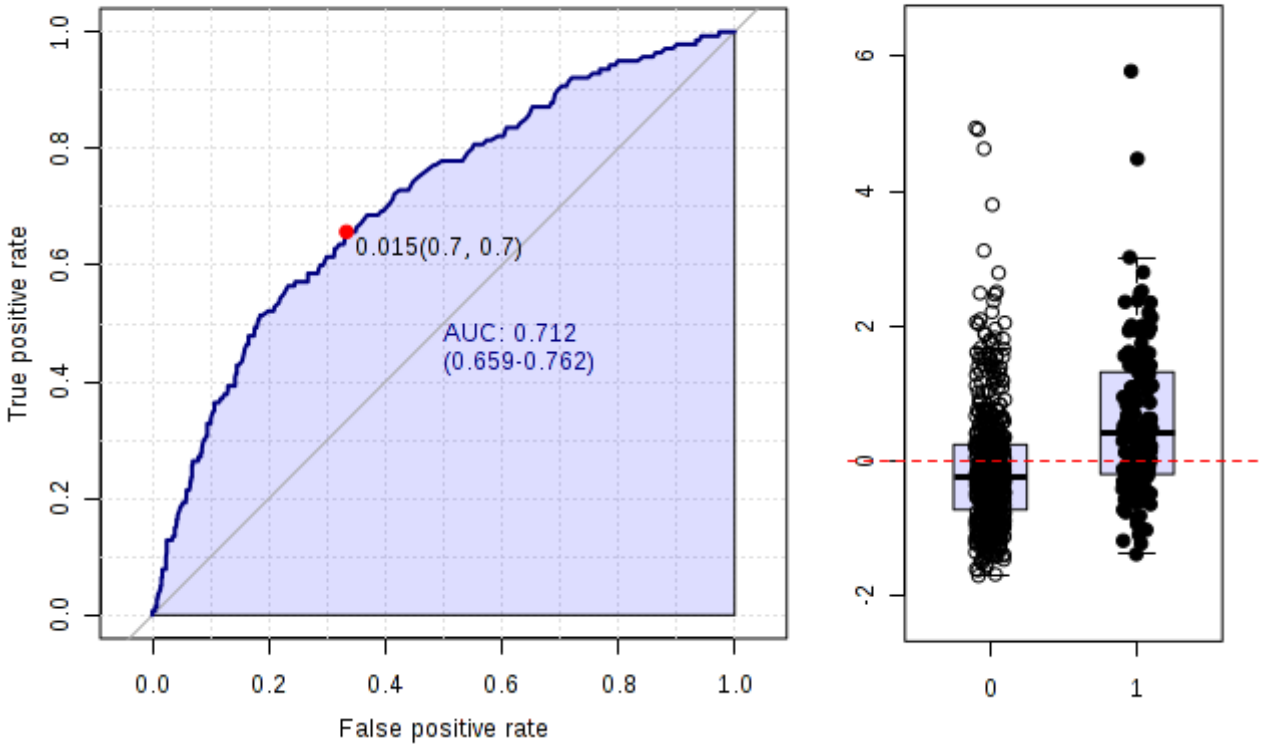


Figure 4.4. ROC curve and boxplot for 2-hydroxybutyric acid; 0-controls; 1-GDM cases.

For comparison, all 26 metabolites significantly associated with GDM after adjustment for confounders and multiple comparison adjustment, were entered into a second ROC curve. The resulting AUC was 0.71 (95% CI: 0.65, 0.76), fractionally less than the AUC using 2-hydroxybutyric acid exclusively, indicating potential over-fitting of the model. The ROC curve results from the metabolite findings were compared with ROC curve results using common clinical risk-factors of GDM used in practice: age, BMI, ethnicity, previous GDM, and family history of GDM. The AUC when using the clinical risk-factors was 0.67 (95% CI: 0.60, 0.73), 0.05 units lower than the AUC when 2-hydroxybutyric acid was used. The addition of the clinical risk-factors and 2-hydroxybutyric acid did not improve the AUC beyond 0.71, which was achieved using 2-hydroxybutyric acid exclusively (Table 4.4).

Table 4.4

Results of ROC Curves Using Different Predictors of GDM

Variables analysed in ROC curve	ROC (95% CI)
All significant metabolites (26)	0.71 (0.65, 0.76)
Single most significant metabolite (2-hydroxybutyric acid)	0.71 (0.66, 0.76)
Clinical risk factors (age, BMI, ethnicity, previous GDM, family history of GDM)	0.67 (0.60, 0.73)
2-hydroxybutyric acid + clinical risk factors	0.71 (0.66, 0.77)

4.4 Discussion

This study was the first to investigate the maternal plasma metabolome in GDM using a large pregnancy cohort. A total of 59 metabolites were significantly different between GDM cases and non-GDM controls following a Mann-Whitney U test, 47 metabolites significantly differed between groups after adjustment for confounding variables (age, body fat [skinfolds], ethnicity, education, and previous GDM), and 26 metabolites retained statistical significance after further adjustment for multiple comparisons. The differences observed in metabolites of significance with and without adjustment for confounders demonstrates a valuable lesson. The removal of significance after adjustment for confounding variables indicates that some of the differences observed between GDM and non-GDM groups reflects already established risk factors. Without adjustment, the information generated would add very little additional value to the field, where gathering demographic data such as BMI is much less time-consuming, less costly, and less invasive using current means, compared with employing a metabolomic technique. This finding highlights the importance of adjusting for confounders, which have been neglected in many previous studies of the GDM metabolome.

The metabolite found to be most significantly increased in GDM in this cohort was 2 - hydroxybutyric acid, an organic acid that results from the degradation of threonine and methionine, and also as a result of the generation of glutathione via the cysteine formation pathway (Landaas & Petterson, 1975). Landaas and Petterson (1975) found that the most important contributor to 2-hydroxybutyric acid production was a high NADH/NAD+ ratio. The AUC resulting from the ROC analysis to determine the ability of 2-hydroxybutyric acid

to identify women who developed GDM was 0.71, higher than the AUC results using the clinical factors currently used to identify women at risk (age, BMI, ethnicity, previous history of GDM, and family history of diabetes), which resulted in an AUC of 0.67. This improvement in the AUC with a single metabolite demonstrates its potential as a biomarker of GDM and warrants further investigation in early pregnancy. There is some criticism that markers of disease determined at the time of diagnosis are not often successfully translated into biomarkers for the prediction of disease. In the case of 2-hydroxybutyric acid, it has been previously found to be an early biomarker of insulin resistance in a non-diabetic population (Gall et al., 2010), which supports its capabilities as a potential early pregnancy biomarker to predict GDM development prior to diagnosis.

A large proportion of the metabolites that remained significant after adjustment for confounders are involved in highly integrated endogenous metabolic pathways. A compound analysis was performed using KEGG online searches (http://www.genome.jp /kegg/compound/) to improve the understanding and interpretation of the study findings. By exploring the inter-relationship between the significant metabolites in this study, hypotheses were able to be drawn regarding the underlying metabolic mechanisms related to GDM. Metabolites that were significantly higher in the GDM participants, compared to non-GDM, were involved in branched-chain amino acid (BCAA) degradation, the tricarboxylic acid (TCA) cycle, use of amino acids for energy production, pathways associated with inflammation, and pathways related to increased reactive oxygen species (ROS). The combination of pathways affected by GDM in this study, and their connection to one another, has led to the hypothesis that dysregulated mitochondrial function, particularly in regard to oxidative phosphorylation, is occurring at the stage of the electron transport chain. The hypothesised dysfunction observed in this cohort is expressed in Figure 4.5.

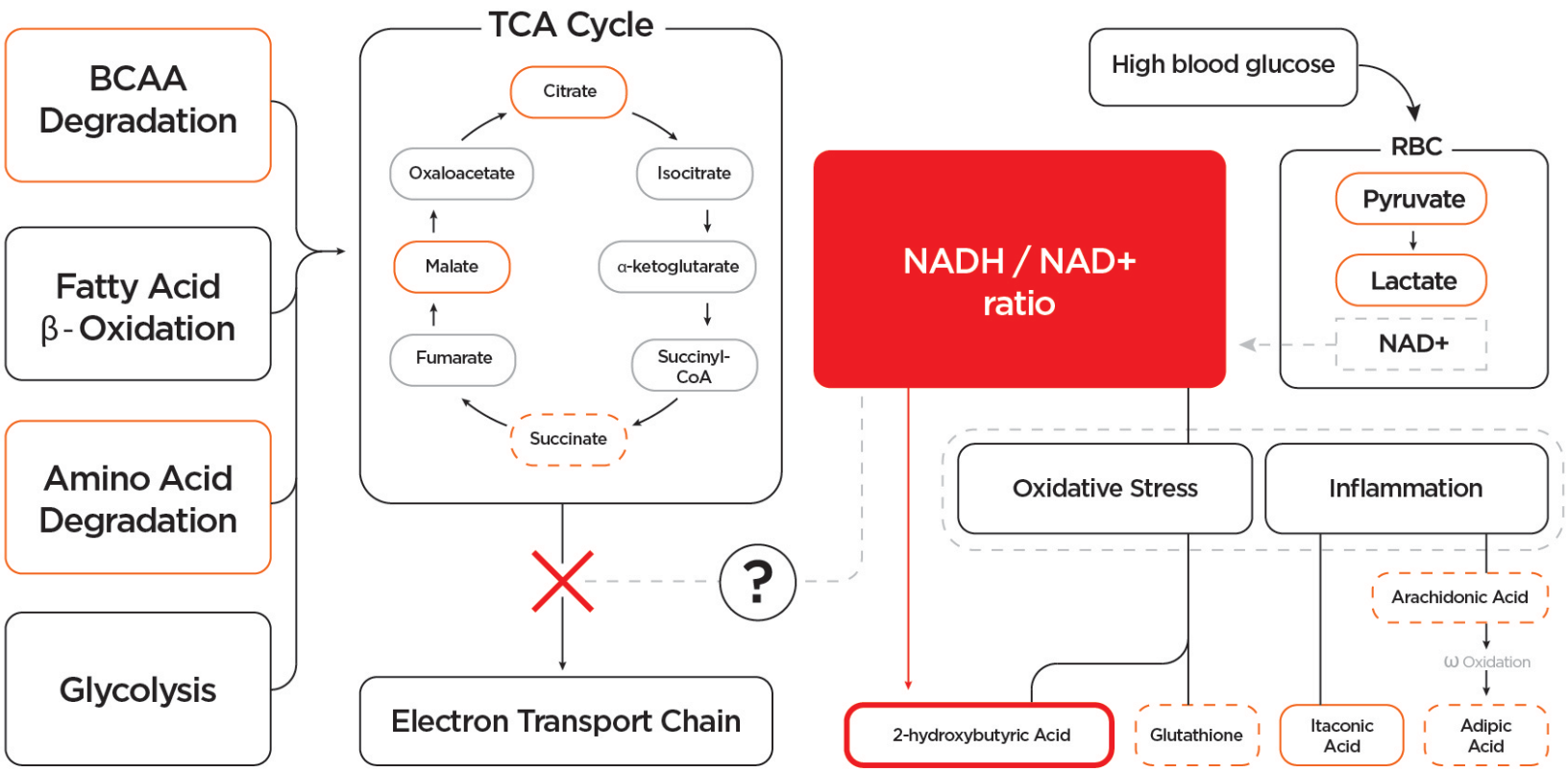


Figure 4.5. Hypothesised mitochondrial dysfunction in GDM related to the electron transport chain.

Solid orange lines represent metabolites or pathways significantly higher in GDM cases after adjustment for confounding variables and multiple comparisons; dashed orange lines indicate metabolites that were significantly higher in GDM after adjustment for confounders ($P < 0.05$), but that did not retain significance after adjustment for multiple comparisons (FDR > 0.05). Features in red indicate the key components of the hypothesis. 2-hydroxybutyric acid is a key component in the hypothesis and is also the most significant metabolite related to GDM in this cohort. BCAA = Branched chain amino acid. TCA = Tricarboxylic acid.

One of the most substantial pathways found to be affected by GDM in this cohort was the degradation of branched-chain amino acids. Higher levels of valine and leucine, along with the breakdown products of the BCAAs, which included 3-methyl-2-oxopentanoic acid, 4-methyl-2-oxopentanoic acid, and 3-hydroxyisobutyric acid, were found in the participants who developed GDM, compared with controls. A number of studies have observed higher levels of BCAAs and their breakdown products in patients with type 2 diabetes (Fiehn et al., 2010; Suhre et al., 2010; Xu et al., 2013; Zheng et al., 2016). In addition, a comprehensive longitudinal study over 12 years, using participants from the Framingham heart study, also found that increased levels of BCAAs in non-diabetic participants were significantly associated with an increased risk of developing diabetes over the course of the following 12 years (T. Wang et al., 2011). Fewer studies have documented this effect in GDM (Enquobahrie et al., 2015; S. Park, Park, Lee, & Kim, 2015), but the metabolic changes associated with GDM development closely parallel those observed in type 2 diabetes. There are multiple hypotheses attempting to explain how BCAAs are related to diabetes. One of the key hypotheses is the reduced activity of the enzyme branched-chain alpha-keto acid dehydrogenase which results in the reduced rates of complete BCAA catabolism and therefore a build-up of BCAAs and their derivatives (Adams, 2011). It has been found previously that a high NADH/NAD+ ratio can contribute to the inhibition of branched-chain alpha-keto acid dehydrogenase (Williamson, Wałajtys-Rode, & Coll, 1979). High levels of 2-hydroxybutyric acid were found in the GDM group in the GUSTO cohort – a metabolite that has been found previously to reflect a high NADH/NAD+ ratio – thereby supporting the hypothesis of reduced activity of the enzyme branched-chain alpha-keto acid dehydrogenase. Results of this study also demonstrated significantly higher TCA cycle intermediates in cases when compared with controls (citrate, malate, and succinate). This observation, in conjunction with a high NADH/NAD+ ratio (reflected by 2-hydroxybutyric acid), has led to the hypothesis of mitochondrial dysfunction, in the form of dysregulated respiration. More specifically, that this build-up of TCA intermediates and available NADH indicates that the electron transport chain is not using the products to generate ATP. This hypothesis provides a metabolic target to consider when exploring therapeutic treatments or prevention strategies. As well as these major pathways of interest, higher levels of metabolites in some pathways associated with inflammation and oxidative stress were also identified (see Figure 4.4). Inflammation and oxidative stress have been reported previously in GDM, and both are also associated with high adiposity, a well-known risk factor for the development of GDM

(Bastard et al., 2006; Furukawa et al., 2004; Wolf et al., 2003). Itaconic acid has been found previously to be higher in women with GDM when compared to healthy controls (de Seymour et al., 2014). Itaconic acid is produced by macrophages in response to inflammation (Michelucci et al., 2013). Arachidonic acid was also observed to be higher in women who developed GDM. Arachidonic acid has been associated with inflammation and has been shown to play a role in the protection against ROS and apoptosis in pancreatic beta cells (Papadimitriou, King, Jones, & Persaud, 2007). It is also a substrate for the production of adipic acid (increased levels in the GDM cases) produced through omega oxidation – an alternative to beta oxidation (Wanders, Komen, & Kemp, 2011). Adipic acid has been found previously in higher amounts in the hair of women in China that developed GDM (X. He et al., 2016). Lastly, a high amount of both pyruvate and lactate in the plasma of women with GDM was observed in this cohort. These metabolites are related to anaerobic glycolysis and the regeneration of NAD+. These findings have led to the hypothesis that one of the major cell types producing additional pyruvate and lactate in response to excess glucose and a high NADH/NAD+ ratio are the red blood cells. Glucose uptake in red blood cells does not solely rely on insulin-dependent transporters and therefore is not overtly affected by insulin resistance in diabetes (Hajjawi, 2013). Red blood cells have actually exhibited an increased glucose uptake when exposed to hyperglycaemia in type 2 diabetes (Bistritzer, Roeder, Hanukoglu, & Levin, 1991). Red blood cells do not contain mitochondria and therefore are limited to anaerobic respiration for glucose utilisation, resulting in pyruvate and its conversion to lactate (van Wijk & van Solinge, 2005).

A limitation of this study was that the plasma samples were collected between the 26th and 28th weeks of gestation – time of GDM diagnosis. Analysing samples collected at time of diagnosis limits the ability to draw firm conclusions around the predictive ability of the identified biomarkers. However, the study provides a robust cohort analysis of the metabolome in GDM, which generated interesting findings that can now be tested in future studies using early pregnancy plasma samples. Identifying metabolites associated with GDM at time of diagnosis informs future studies as to which metabolites might be beneficial to quantitatively investigate. In addition to finding individual metabolites that would warrant further investigation as biomarkers, characterising the metabolome as a collection of highly interrelated metabolites assisted the discovery of information important for targeting prevention/therapeutic strategies to re-route metabolic trajectories that would lead to the dysfunction observed in GDM.

The in-house metabolite library of MCF derivatised standards used in the study is limited to 217 compounds, and therefore many peaks observed in the total ion chromatograms remained undetected using the software employed. Using the NIST14 library partially solved this limitation, but comes with its own limitations. The NIST14 library only contains information on the mass of compounds, not on expected retention time. This means that a proportion of the metabolites detected in this study using the NIST14 library only have low match percentage with the library match. The certainty of the match can only be confirmed using standards that are subjected to the same procedure/protocols employed, in order to fully confirm their identification. This is a costly and time-consuming task.

One of the unavoidable limitations of this study is that plasma is a complex biofluid. It reflects overall metabolism, being derived from a range of different tissues in the human body, meaning that it is not possible to confirm where the metabolites originate from in an untargeted metabolomics study. The purpose of an untargeted metabolomics study is to explore differences and build hypotheses. As the majority of the significant metabolites are endogenous in origin, it has led to the hypothesis presented, relating to a dysregulation in mitochondrial function – a hypothesis that has been drawn based on the combination of the study findings and supporting literature, which makes the biological hypothesis plausible. More recently however, there has been a surge in research around the gut microbiome, and its interaction with human health and disease. Given that previous research has found an association between the gut microbiome, insulin resistance in a non-diabetic population, and the serum metabolome in that population, it is essential that the potential effect of the microbiome on the plasma metabolome is not ignored in the current study (Pedersen et al., 2016). The current study was not designed for the purpose of exploring the association between the microbiome and GDM, and as such, the gut microbiome has not been characterised in these participants. Therefore, it is difficult to draw any conclusions on the relationship between the gut microbiome and the plasma metabolomic findings in this study. It is worth noting however that Pedersen et al. (2016) found that BCAA metabolism was highly correlated with the gut microbiome, particularly with the species *P. copri* and *B. vulgatus*. Research has also shown that probiotic supplementation during pregnancy resulted in a significantly reduced rate of GDM in the treatment group (supplemented with *Lactobacillus rhamnosus GG* and *Bifidobacterium lactis Bb12* at a dose of 10^{10} colony-forming units/d each) compared to the placebo-controlled group (Luoto, et al., 2010; 13% vs

36 %), demonstrating that the microbiome is clearly an important facet of GDM development that needs to be considered in future studies.

The greatest strength of this study was access to a large pregnancy cohort which had performed universal screening of GDM in all participants, collected biological samples, and recorded important demographic variables, allowing for the ability to account for the effect of major confounders. In addition, the cohort consisted of a sample of women from a population that is at high risk for GDM, as reflected in the high rate of GDM diagnosed in the cohort (17.8% of women were diagnosed with GDM using the WHO criteria for diagnosis, which is a relatively conservative measure, compared to other criteria that are currently employed). There is also much less research conducted in Asian populations, compared with their Western counterparts, so these findings add great value to understanding the metabolic effects of GDM on this high-risk group.

4.5 Conclusion

The study conducted has identified a biomarker of GDM (2-hydroxybutyric acid) which has been found previously to be associated with insulin resistance and dysregulated glucose metabolism. A set of 47 metabolites were also identified that demonstrated a significant positive association with GDM (except for one metabolite demonstrating a negative association), after adjustment for confounding variables. A collection of these metabolites is closely linked in endogenous metabolic pathways. Future studies should investigate the metabolites identified in this cohort as significantly associated with GDM, in early pregnancy, to test their ability as biomarkers for detecting women at risk of GDM development. Pathway analysis of the significant metabolites led to the hypothesis that the women in the Singapore cohort who developed GDM experienced dysregulated mitochondrial function at the stage of the electron transport chain. It was also observed that higher amounts of some metabolites in the participants who developed GDM, when compared to non-GDM, were linked with oxidative stress and inflammation. The metabolic pathways of interest could be used to drive investigations into prevention strategies for GDM. One way to achieve this is to investigate the crossover between the GDM metabolome and the effects of modifiable lifestyle factors, such as maternal diet, on the metabolome. The point of intersect between the GDM metabolome and the diet-related metabolome would be the starting point for future investigations into *how* the maternal diet may be related to GDM development.

Chapter Five: The Hair Metabolome in GDM

5.1 Introduction

Studies of the GDM metabolome have provided metabolic information that could be useful for the identification of potential biomarkers and metabolic mechanisms underlying GDM development. However, thus far, studies of the GDM metabolome have predominantly been conducted on small sample sizes (see Table 1.6), with insufficient power to detect robust, reproducible differences between case and control groups. Additionally, there has been substantial inconsistency in the study findings. There are multiple possibilities for this, the highest contributor likely to be the lack of standardisation between metabolomics studies – using different analytical platforms, sample processing protocols, specimen types, normalisation procedures, and statistical analysis approaches. Most research conducted on the GDM-related metabolome has consisted of studies using blood, urine, or amniotic fluid. Urine is highly influenced by recent dietary exposure, the gut microbiome, and pharmaceutical use, which can complicate the interpretation of findings (Bouatra et al., 2013; Rasmussen et al., 2011). Plasma studies have produced some interesting results, but have yet to identify clinically useful biomarkers of GDM development. A possible explanation for this is the lack of studies conducted using large, representative cohorts (addressed in Chapter 4 of this thesis). One pilot study has used the hair metabolome for differentiating women who developed GDM, from healthy controls, identifying one metabolite that was significantly different between GDM cases and non-GDM controls (X. He et al., 2016).

The hair metabolome potentially provides a longer-term snapshot of the metabolic state of an individual than biofluids such as plasma and urine. The lifecycle of an erythrocyte is approximately 115 days (Franco, 2012) and urine is regenerated multiple times throughout a day (Saude, Adamko, Rowe, Marrie, & Sykes, 2007). In comparison, hair grows at a rate of ~1 cm per month (Harkey, 1993). Therefore, a strand of hair 9 cm long could provide information about the metabolic state of a woman averaged throughout the course of her pregnancy (assuming the hair strand was collected from the end closest to the scalp), or, if carefully segmented, could provide information specific to trimesters. Human hair composition is predominantly protein (65-95%), but also includes other fatty acids, complex lipids, polysaccharides, water, and other trace elements (Horvath, 2009). In addition to the benefit of capturing longer-term information, compared to other biofluids, the hair metabolome is non-invasive to collect, and requires low-maintenance storage. Hair has been

recovered from archaeological sites dating back as far as 195,000 years (Backwell et al., 2009), supporting the concept that hair is well preserved, even in uncontrolled conditions. The non-invasive collection and low-maintenance storage of hair makes it a favourable biological specimen in low-resource settings such as developing countries where it is usually very difficult to set up collection (and appropriate storage e.g., −80 ºC freezers) of other specimens such as blood, tissue, or amniotic fluid. Hair samples are also much safer to collect. The non-invasive collection process poses little risk to the participant or the patient, and hair, when compared to other fluids such as blood, poses less risk of transmissible disease to the research team collecting and analysing specimens (Heymann & American Public Health Association, 2008). Despite hair being a more robust and easy to collect sample, collection of hair does require close adherence to a standard operating procedure, to ensure that samples are collected in a standardised manner, and to facilitate an accurate comparison between sample groups. Hair samples should be taken from the same area of the head from each participant, and the distance between the scalp and collection point needs to be consistent.

Human hair has been utilised historically to detect drug exposure (Graham, Koren, Klein, Schneiderman, & Greenwald, 1989) and mineral status (Deeming & Weber, 1978). More recently, hair has been analysed to detect environmental toxins (Covaci, Tutudaki, Tsatsakis, & Schepens, 2002), detection of toxins from dietary consumption (Schaefer, Jensen, Bossart, & Reif, 2014), and even to differentiate dietary pattern consumption (Cho et al., 2011; Patel et al., 2014; Valenzuela et al., 2012). Much of the work conducted in hair has been performed in the field of forensic science. One study to date has investigated the hair metabolome related to GDM. This small case-control study conducted in Chongqing, China, found that adipic acid was significantly higher in GDM cases when compared to controls (X. He et al., 2016). The authors hypothesised that the discovery of higher levels of adipic acid in GDM implied dysregulated fatty acid oxidation. This is a plausible explanation given the increased circulating free fatty acids found in women with GDM, which can activate the omega oxidation pathway in an attempt to reduce the levels of available fatty acids (Meyer, Calvert, & Moses, 1996). The pilot study by X. He et al. (2016) demonstrated that the hair metabolome can act as a source of metabolic information for differentiating women who develop GDM from those with a healthy pregnancy outcome.

The objective of this study was to investigate the hair metabolome in pregnancy, with the aim of identifying a metabolomic signature of GDM, using participants from a large multi-ethnic Asian cohort - GUSTO.

5.2 Methods

5.2.1 Participants.

The hair metabolome was characterised for a subset of participants from GUSTO. The details of the cohort and associated ethics approval for this work have been described in Chapter 2 of this thesis.

GUSTO participants were eligible for this study of the hair metabolome if they met the following criteria:

d) underwent an (OGTT) at 26-28 weeks' gestation;

e) had at least three strands of maternal hair available in the biobank;

f) had complete data for maternal age, BMI, and ethnicity (as these are three major clinical predictors of GDM in pregnancy).

Missing values for discrete demographic variables were imputed using a multiple imputation procedure (Rubin, 1996) available in SPSS version 23.0 for birth order (N = 10), education level (N = 8), monthly household income (N = 49), family history of diabetes (N = 28), previous history of GDM (N = 19), smoking (N = 1), and alcohol consumption (N = 20).

5.2.2 Hair collection.

Hair was collected from participants at the time of the OGTT (between the 26th and 28th weeks of gestation) following a SOP (Appendix 8). Five strands of hair were cut 0.5 cm from the scalp at the occipital region of the back of the head. The strands were then contained in an air-tight zip-lock bag and stored at −20 ºC in Singapore, until required for analysis.

5.2.3 Metabolomic profiling of hair.

For this study, three strands of hair from each participant were shipped from Singapore to New Zealand in January, 2015. Samples were randomised prior to extraction according to a stratified randomisation sequence provided by a biostatistician, in order to prevent bias from technical variation. The stratified randomisation ensured that cases of GDM were spread equally throughout the 57 days of metabolomic analyses. The identity of the samples was

blinded to researchers until statistical analysis, to reduce unintentional bias. Samples were stored at −20 ºC in New Zealand until analysis.

5.2.3.1 Hair preparation and internal standard addition

Four internal standards (Sigma-Aldrich) were prepared for addition to hair samples prior to the experimental processing; alanine-d_4 (99%, 10 mM), phenylalanine-d_5 (99%, 10 mM), tyrosine-d_2 (99%, 10 mM), and citric acid-d_4 (99%, 10 mM). Alanine-d_4, tyrosine-d_2, phenylalanine- d_5, and citric acid- d_4 were dissolved separately into 10 mL of milli-Q water and then combined to form an internal standard (IS) mix of 40 mL in total.

The metabolite extraction method was adapted from that published by X. He et al. (2016) and Sulek et al. (2014). The derivatisation procedure followed was adapted from Smart et al. (2010).

The hair strands underwent two cycles of a washing protocol to remove any exogenous compounds from the outside of the hair (e.g., hair gels/sprays/shampoo residue). Each wash cycle consisted of a 30-s vortex with 20 mL of milli-Q water, followed by a 30-s vortex with 20 mL of 100 % methanol (LC-MS grade, Merck). The samples were left to dry in a fume hood. Once dry, each participant's hair strands were cut into 5 mm segments, and the hair was weighed out and recorded for use in biomass normalisation of the data. The samples were placed in a tared glass vial. The hair then underwent hydrolysis in a concentrated alkaline solution to break down the physical structure of the hair strands and access the endogenous metabolite information. The alkaline hydrolysis was performed by adding 400 µL of potassium hydroxide (1 M, Merck) to each of the glass vials, along with 40 µL of the prepared IS mix, for two incubation periods. The first incubation period was 30 min at 54 °C. After the first incubation, the samples were centrifuged at 4000 rpm for 5 min at room temperature (Eppendorf 5430R, Mediray, New Zealand), to ensure that all hair fragments were fully submerged in the solution prior to the second incubation. The second incubation period was 18 hr at 54 °C.

5.2.3.2 Neutralisation and extraction

The base-hydrolysed hair samples were neutralised to pH 7 by addition of 67 µL of sulfuric acid (3 M, ROMIL Pure Chemistry). Salts and proteins from the hair were precipitated by addition of 1 mL of methanol. The sample was centrifuged at 4000 rpm for 5 min at room temperature (Eppendorf 5430R, Mediray, New Zealand), to separate the precipitate from the supernatant. The supernatant was transferred into three microcentrifuge tubes, 350 µL into

each tube. One tube was used for analysis, and the other two were stored at −80 °C for future analyses by other research groups. Any remaining supernatant was collected into a falcon tube placed on dry ice throughout the experiment, to form a pooled solution from which to make QC samples. The 350 µL extract was evaporated to dryness using a SpeedVac concentrator (Savant SC250EXP, ThermoFisher Scientific) with a refrigerated vapor trap (Savant RVT4104, ThermoFisher Scientific) at 0.8 hPa, for 18 hr. Once dry, samples were placed in the −80 °C freezer until further analysis. Once all samples had been extracted, the pooled QC mix was defrosted in a cold room (4 ºC) and separated into 350 µl aliquots. The aliquots were evaporated to dryness as per the samples, then placed into the −80 ºC storage prior to analysis.

5.2.3.3 Methyl chloroformate (MCF) derivatisation procedure.

The derivatisation procedure employed was first published by Smart et al. (2010), and was used successfully by X. He et al. (2016), and Sulek et al. (2014) in their studies of the hair metabolome.

In brief, dried samples and QCs were resuspended in 200 µL of 1 M NaOH, vortexed for 20 s and placed on wet ice for 20 min before being further vortexed for 10 s. The resuspended sample was transferred to silanised glass test tubes. Each sample had the addition of 167 µL of methanol and 34 µL of pyridine, followed by a 10-s vortex. To begin the alkylation reaction, 20 µL of MCF was added to the sample and the mixture was vortexed vigorously for 30 s. Immediately following the vortex, another 20 µL of MCF was added, followed by a further 30-s vortex. While the sample was still on the vortex, 400 µL of chloroform was added and vortexing continued for a further 10 s. Finally, 400 µL of 50 mM sodium bicarbonate was added to complete the reaction, and was vortexed for a final 10 s. Derivatised samples were centrifuged for 3 min at 2000 rpm to separate the aqueous and organic phases (Eppendorf 5430R, Mediray, New Zealand). The upper phase (aqueous) was removed and removal of any residual water in the organic phase was performed by addition of 300 mg of anhydrous sodium sulfate. The organic layer was then transferred from the silanised tube to a glass gas chromatography-mass spectrometry (GC-MS) vial with an insert, for GC-MS analysis.

5.2.3.4 Gas chromatography-mass spectrometry analysis.

Each sample was analysed using GC-MS for the detection of amino acids, fatty acids, and other organic acids, as per the plasma metabolome analysis (Section 4.2.3.4). The 10-min

bake performed at the end of the run in the plasma analysis was reduced to 2 min in the hair metabolome analysis as there were no late-eluting compounds from the hair samples, reducing the likelihood of carryover. The total time of the GC-MS run for each sample before cooling, was 40 min.

5.2.4 Data processing and cleaning

Data processing from raw GC-MS files, and manual checking and correction prior to normalisation, were conducted using the same processes followed for the plasma metabolomic data (Sections 4.2.4 and 4.2.5).

5.2.5 Statistical analysis

5.2.5.1 Normalisation of metabolites and removal of outliers.

The same normalisation strategies that were used in Section 4.2.6.1 for the plasma metabolites, were employed for normalisation of the hair metabolites. As per the processes described, the metabolites underwent a median-centering batch normalisation, followed by normalisation with the internal standards (only if the metabolite had a correlation with one of the internal standards $R > 0.9$ in the QC samples). Unlike the internal standard normalisation performed in the plasma dataset, which resulted in internal standard normalisation of 130 metabolites, only 42 metabolites met the criteria for internal standard normalisation in the hair dataset. This implies that the internal standards were not behaving similarly to many of the metabolites detected in the hair QC samples.

5.2.5.2 Missing values and outliers.

The final data contained 0.7% missing values. Missing values were imputed as half of the minimum value detected in the samples, an approach previously reported by Xia et al. (2009). The metabolite scores were standardised to Z-scores and a PCA was performed using all hair metabolites, to identify outliers. Outliers were removed so that they did not heavily influence the downstream analyses.

5.2.5.3 Analysis of metabolite differences between groups.

The two approaches used to test for significant differences between the GDM cases and non-GDM controls of the metabolites in the plasma metabolome (Section 4.2.6.3) were also used for comparing differences in hair metabolites between groups. Receiver operating characteristic (ROC) curves were generated using Metaboanalyst V 3.0 for any hair

metabolites found to be significantly associated with GDM. All analyses were performed using SPSS v 23.0. A P-value < 0.05 was considered statistically significant.

5.3 Results

5.3.1 Participants

A total of 733 women met the inclusion criteria listed in Section 5.2. After removal of outliers according to the visual inspection of the principal component analysis plot (Figure 5.1), a total of 728 participants remained in the study for further analysis. The characteristics of the participants in the study are listed in Table 5.1. Participants were from three ethnic groups: Chinese (55.5%), Malay (27.3%), and Indian (17.2%). The mean (± standard deviation) age of the participants was 30 (± 5) years and median (lower quartile, upper quartile) BMI was 25.3 (23.0, 28.2). Of 728 participants: 17.3 % (N = 126) were diagnosed with GDM according to the WHO criteria (Alberti & Zimmet, 1998). Women diagnosed with GDM were more likely to have: higher BMI and body fat, advanced age, a history of GDM in a previous pregnancy, attained a higher education, and reported a higher household income level.

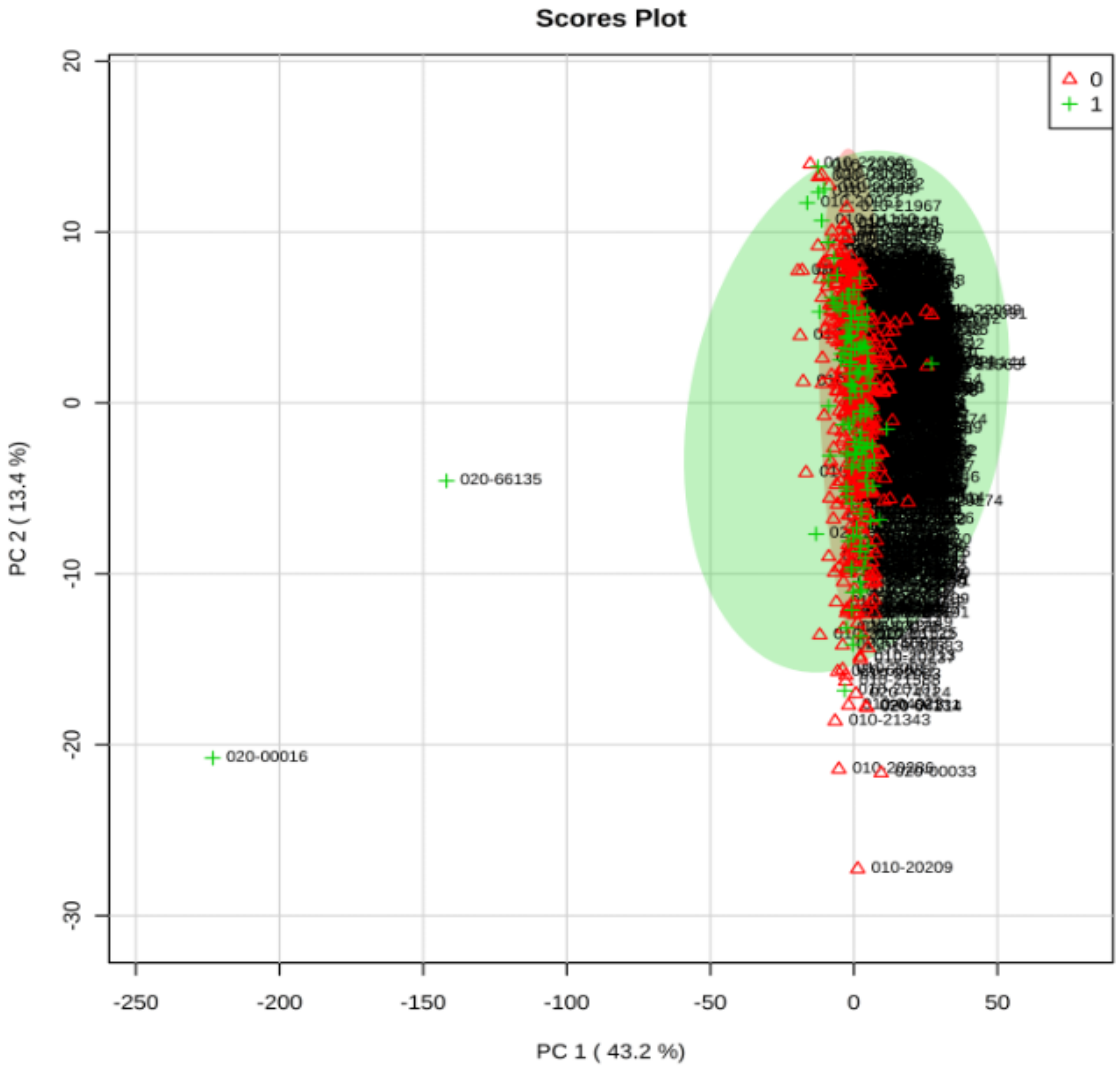


Figure 5.1. Principal component analysis of participants according to their complete hair metabolite profile revealed five outliers.

Table 5.1

Participant Characteristics

		Gestational Diabetes Mellitus			
		No (N = 602 [82.7 %])		**Yes (N =126 [17.3 %])**	**P-Value**
BMI (median [lower quartile, upper quartile])		25.0 (22.9, 27.9)		26.2 (23.8, 29.3)	0.02*
Sum of skinfolds (mean ± SD)		106.7 (22.4)		114.5 (22.4)	< 0.001*
Age (mean ± SD)		30 ± 5		32 ± 5	< 0.001*
Ethnicity	Chinese	320 (53.2)	Chinese	84 (66.7)	
N (%)	Malay	185 (30.7)	Malay	14 (11.1)	
	Indian	97 (16.1)	Indian	28 (22.2)	< 0.001*
Education	< Secondary	203 (34.2)	< Secondary	24 (19.0)	
N (%)	Post-Secondary	213 (35.9)	Post-Secondary	39 (31.0)	
	University	177 (29.8)	University	63 (53.0)	< 0.001*
Household monthly income	< 2000	86 (16.0)	< 2000	10 (9.0)	
(SGD)	2000-6000	304 (56.6)	2000-6000	57 (51.4)	
N (%)	> 6000	147 (27.4)	> 6000	44 (39.6)	0.02*
Smoking	No	582 (96.7)	No	123 (98.4)	
N (%)	Yes	20 (3.3)	Yes	2 (1.6)	0.40

		Gestational Diabetes Mellitus			
		No (N = 602 [82.7 %])		**Yes (N =126 [17.3 %])**	**P-Value**
Alcohol consumption	No	572 (97.8)	No	123 (99.2)	
N (%)	Yes	13 (2.2)	Yes	1 (0.8)	0.48
Previous history of GDM	No	564 (97.9)	No	109 (87.9)	
N (%)	Yes	12 (2.1)	Yes	15 (12.1)	< 0.001*
Family history of diabetes	No	406 (70.5)	No	80 (64.5)	
N (%)	Yes	170 (29.5)	Yes	44 (35.5)	0.20
Birth order	First child	256 (43.2)	First child	46 (36.5)	
N (%)	Not first child	337 (56.8)	Not first child	80 (63.5)	0.20
Birthweight (g) (median [lower quartile, upper quartile])		3090 (2860, 3386)		3120 (2850, 3340)	0.67
Gestational age at delivery (days) (median [lower quartile, upper quartile])		272 (266, 278)		271 (264, 277)	0.03*

Chi-square (2-tailed) tests were conducted to compare group differences of categorical variables, Independent t-test was conducted for continuous variables that followed a normal distribution, and Mann-Whitney U test was conducted for continuous variables that did not follow a normal distribution

*P < 0.05

5.3.2 Metabolites identified.

A total of 277 metabolites were identified in the hair metabolome of the participants. Of the 277 compounds identified, 40.8% were identified (80-100% match), 25.6% were putatively identified (60-79% match), and 33.6% were unknown (< 60% match). The most common classes of compounds identified or putatively identified were amino acids or derivatives (23.9%), fatty acids or derivatives (10.9%), carboxylic acids (9.8%), benzenoids (9.2%), and peptides (7.6%). The list of identified metabolites, the metabolite library they were derived from, and their match factor (%) with the library standard, are listed in Appendix 9.

5.3.3 Significant differences in metabolites between groups.

A Mann-Whitney U test to compare differences in median metabolite levels between GDM cases and non-GDM controls resulted in 25 metabolites of significance ($P < 0.05$). After adjustment for multiple comparisons using a Benjamini-Hochberg technique (Benjamini & Hochberg, 1995), none of the metabolites remained statistically significant. The false discovery rate (FDR) for the 25 metabolites with $P < 0.05$ was 0.50 (Table 5.2).

Table 5.2

Metabolites Detected as Significantly Different Between Cases and Controls Using a Mann-Whitney U Test

Metabolite	**P-Value**	**False Discovery Rate (FDR)**
2,4-Di-tert-butylphenol	0.008	0.50
Tentative Ethanedioic acid, dihydrazide	0.01	0.50
2-Aminobutyric acid	0.01	0.50
Alanine	0.01	0.50
2-Methyl-1,6-dihydro-4-methylamino-6-pyrimidinone	0.02	0.50
Decanoic acid	0.02	0.50
N,4-Diethyl-4-heptanamine	0.03	0.50
Phthalic acid, 2-ethoxyethyl ethyl ester	0.03	0.50
1-Methylpyridin(2H)-2-one-5-carboxylic acid, methyl ester	0.03	0.50
4-Amino-2,6-dihydroxypyrimidine	0.03	0.50
Unknown 201(100).117.(89.6).119(86.2)	0.03	0.50
Tentative Diethyl 2,2'-(2,2'-oxybis(ethane-2,1-diyl)bis(oxy)) diacetate	0.03	0.50
1H-Indole, 3-[(4,5-dihydro-3,5,5-trimethyl-1H-pyrazol-1-yl)methyl]-	0.03	0.50
cis-1,3-Diacetamido-1,2,3,4-tetrahydronaphthalene	0.03	0.50
p-Chloroaniline	0.03	0.50
Unknown 074(100).46(21.3).59(0.97)	0.04	0.50

Metabolite	P-Value	False Discovery Rate (FDR)
Glutamic acid	0.04	0.50
Morpholin-4-yl-acetic acid, hydrazide	0.04	0.50
l-Norvaline, N-methoxycarbonyl-, hexyl ester	0.04	0.50
Unknown 135(100).180(28.0).77(19.9)	0.04	0.50
Ethanone, 1,1'-(1,4-dihydro-2,4,6-trimethyl-3,5-pyridinediyl)) bis-	0.04	0.50
Benz[d]isoxazole-5-ol-4-one, 4,5,6,7-tetrahydro-3-methyl-	0.05	0.50
Glycine, N-methyl-n-propoxycarbonyl-, ethyl ester	0.05	0.50
Glutathione	0.05	0.50
Unknown 070(100).239(38.0).128(13.0)	0.05	0.50

The results of the logistic regression, after adjustment for confounders demonstrated a significant association between the relative abundance of alanine and GDM ($P < 0.05$; Table 5.3), with lower levels in women who developed GDM. However, adjustment for multiple comparisons indicated an FDR of 0.99, deeming it extremely likely to be a false positive result of multiple testing, rather than a true biological difference between cases and controls.

Table 5.3

Metabolites Detected as Significantly Associated with GDM After the Adjustment for Confounding Variables

Metabolite	Odds Ratio	95% CI	P-Value	FDR	Fold Change (healthy/GDM)	Coefficient of variation
Alanine	0.79	(0.64, 0.99)	0.04	0.99	1.46	0.06

Confounding variables adjusted in logistic regression model: age, sum of skinfolds, ethnicity, education, previous GDM

5.3.4 Receiver operating characteristic curves.

Alanine, the metabolite in the hair metabolome that was the most significantly associated with GDM after adjustment for confounding variables, was entered into a ROC curve to test its ability to correctly classify GDM cases from non-GDM in this cohort. The resulting AUC (95% CI) was 0.57 (0.52, 0.63) (Figure 5.2), too low to be deemed a clinically useful biomarker.

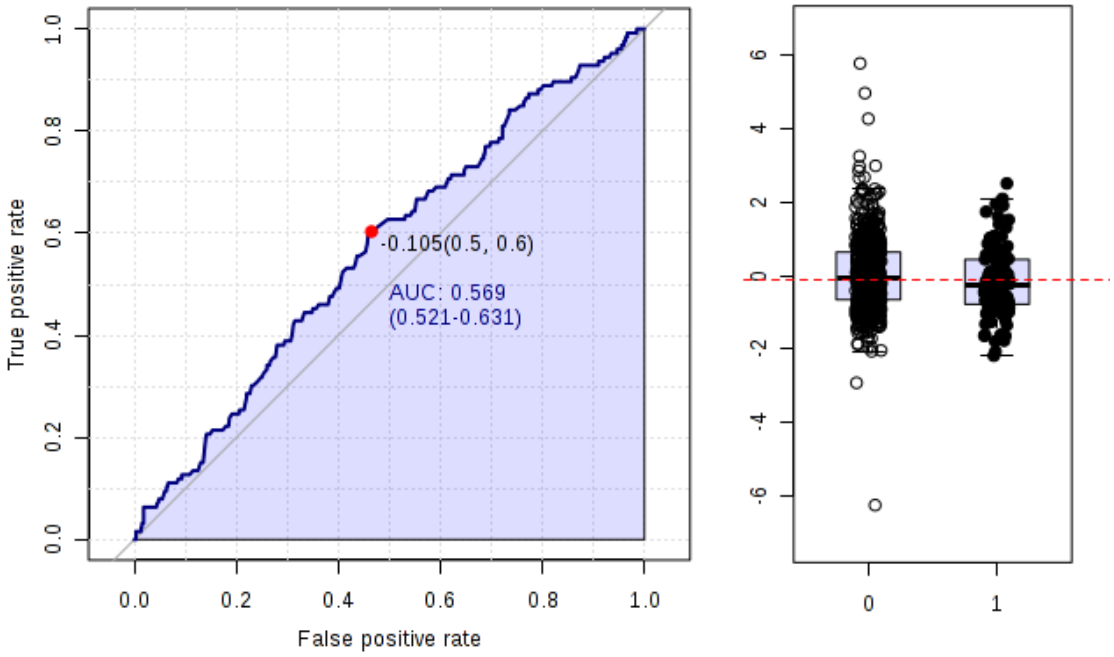


Figure 5.2. ROC curve and boxplot for alanine from the hair metabolome as a biomarker of GDM.

Note 0-controls; 1-GDM cases

5.4 Discussion

This study was the second to investigate the hair metabolome related to GDM, following a small pilot study that demonstrated the potential of the hair metabolome to differentiate women who did and did not develop GDM. After the adjustment for confounding variables, only alanine was significantly associated with GDM development ($P < 0.05$) in this cohort, with significantly lower levels found in women who developed GDM, compared to those that did not develop GDM. However, after the adjustment for multiple comparison testing, the FDR was close to 1, indicating a very high likelihood that this finding was a false positive.

The resulting conclusion was that using the method described in Section 5.2 for the metabolomic analysis of the hair metabolome, no meaningful differences were observed between women who developed GDM and those who did not. These findings are in contrast with the findings of the pilot study conducted by He et al. (2016). The small case-control study found that adipic acid was significantly higher in GDM cases when compared to controls. In the current study, adipic acid was not found to be significantly associated with GDM, even prior to adjustment for multiple comparisons and confounding variables. The metabolomic methods employed were similar between studies, and samples were collected at the same time (at the time of OGTT). The biggest difference between the two studies was the number of participants. The pilot study was conducted using a total of 94 pregnant women from Chongqing, China, whereas the current study was conducted on a subset of women from a large Singaporean pregnancy cohort, a total of 728 participants. Studies have raised concerns over the susceptibility of small studies to strong random effects which can result in false positive significant findings being reported, which commonly disappear when validation is carried out on a larger group (Button et al., 2013; Ioannidis, 2005). Furthermore, there were significant differences in the hair biomass of the participants between the two studies. In the former study, 10 mg of hair was available from each participant, whereas in the current study, the sample biomass ranged from 0.5 mg to 5.5 mg, significantly less biomass by comparison. This difference in biomass may explain the differences in findings. In addition, the current study used a multi-ethnic Asian cohort from Singapore, consisting of Chinese (56%), Malay (27%), and Indian (17%) participants, whereas the pilot study contained only Chinese participants. It was plausible that the hair metabolome in GDM of different ethnic groups may be significantly different, reducing the appropriateness of direct comparisons between the two studies. To test this hypothesis, a Mann-Whitney U test was performed on adipic acid between GDM cases and non-GDM controls, after stratification by ethnicity. Adipic acid was not significantly associated with GDM in any of the three ethnic groups ($P > 0.05$), including the Chinese participants. Therefore, it is less likely that this inconsistency is due to ethnicity, and more likely it may be associated with differences in biomass analysed.

The most apparent limitation of this study was the limited biological specimens available from each participant – as low as 0.5 mg of hair. In addition, there was large inter-individual variability in the biomass between individuals in the study (0.5 - 5.5 mg). Despite correction for hair biomass in the normalisation strategy, it is possible that the variation in biomass is not fully accounted for, as the response in some metabolites may not be linearly associated

with biomass. During the course of the experiment, visual inspections were conducted using the total ion chromatograms produced by the GC-MS runs. There was no clear indication that low biomass compromised the quality of the analysis (see examples in Figure 5.3). Future studies should test the response of the metabolites across different biomass values to determine if new normalisation strategies for biomass adjustment need to be developed.

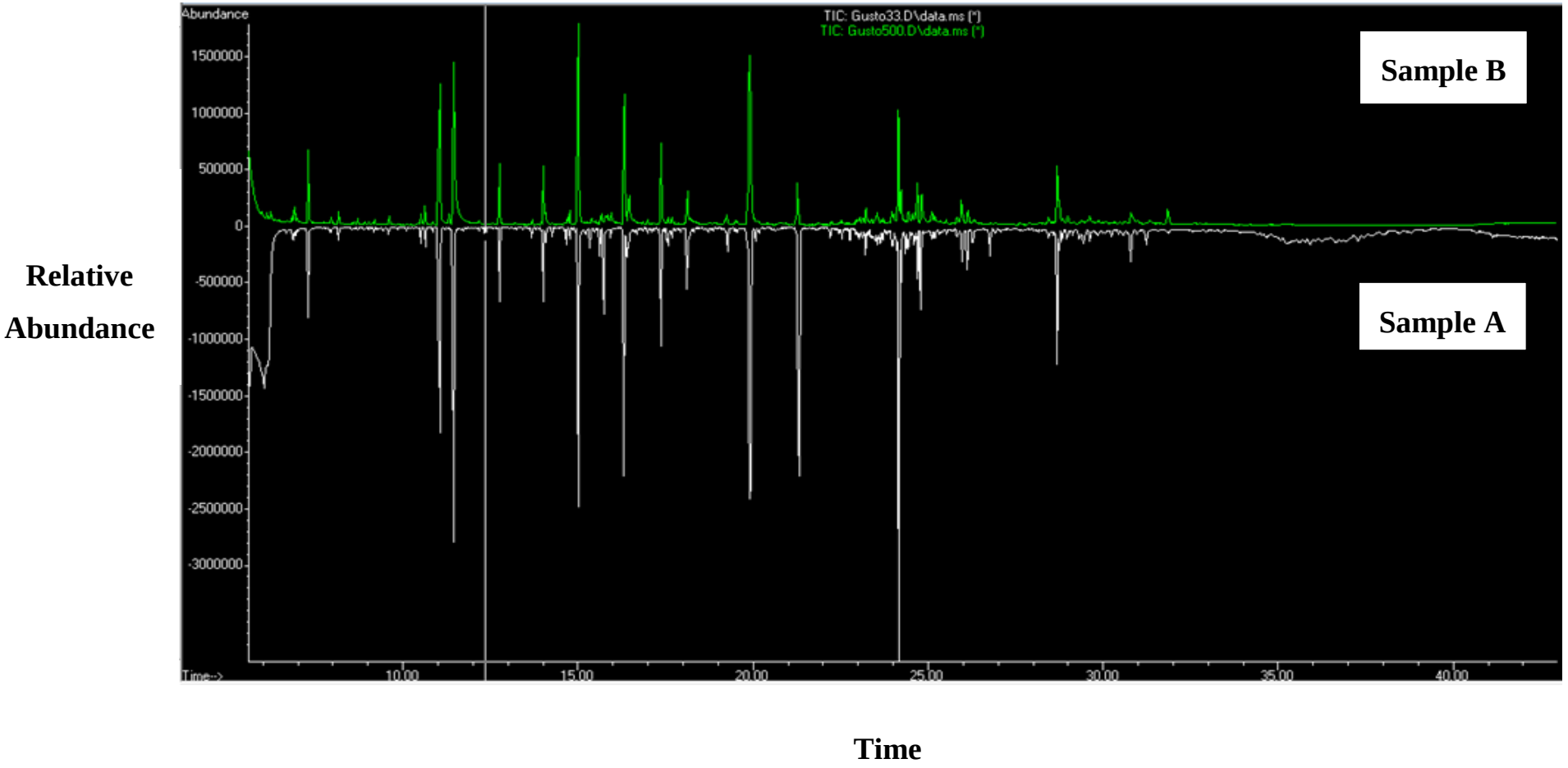


Figure 5.3. A comparison of the total ion chromatogram from GC-MS analysis of two hair samples with different biomass shows that the quality of the chromatogram was not compromised due to the low biomass. Sample A has a hair biomass of 0.5 mg; Sample B has a hair biomass of 5.5 mg.

As a result of limited biomass availability, it was necessary to utilise the full amount of sample provided. As such, there was not sufficient biomass to explore hair segmentation in this study. Segmentation of the hair is a preferred method as it would have allowed for analysis at a targeted time-frame of exposure. For example, if the 9 cm closest to the scalp was analysed then it would have been possible to look specifically at the pregnancy time point (provided all participants had at least 9 cm of hair length). This is something that needs to be considered early in the design of future studies, when specimen collection protocols are being developed. This will also require knowledge of which end of the collected hair sample was from the scalp, which can be achieved using a marking at time of collection, or through microscope investigation of the hair strands at the time of analysis. An additional limitation, related to the biomass restrictions in this study, was that not all participants had the same hair length. The inter-individual difference in hair length and how it might affect analysis is not something that had been considered previously in studies of the hair metabolome. The effect of hair length is important as it represents the time-frame being analysed. Therefore, the longer the hair, the more dilute the results would have been (assuming the greatest difference between GDM cases and non-GDM controls was during pregnancy). In future studies, if it is necessary to use full hair strands due to constraints on biomass, the hair strand length should be measured and recorded to allow for adjustments post-analysis.

The major strength of this study was that it was the first to investigate the hair metabolome in a large pregnancy cohort. In theory, analysing a large, representative sample group provides the power to identify subtle changes in the metabolome. A cohort study also reduces the risk of inflating some differences which can occur in case-control studies. In case-control studies, where the ratio of cases to controls does not adequately represent the true prevalence of the disease, this can bias the results by not encapsulating the full extent of variation in controls. Analysing hair from a large cohort produces results that are more relevant and translatable to the population of interest. The disadvantage of working with a large number of samples, however, is that analysis takes much longer. This study required the GC-MS to be operated for 57 days consecutively, which introduced new challenges that are not often faced (or not to such a high degree) by many researchers using mass spectrometry. The biggest challenge was dealing with the introduction of technical/machine variation over time. Multiple methods were tested to correct this variation (described in Section 4.2.6.1) before selecting a normalisation strategy that effectively removed the technical variation (as determined by a Kruskal-Wallis test across batches). The strategy that was able to achieve the removal of

batch effect in this study was median centring using the samples. This approach is considerably more conservative compared to other methods, but was performed to ensure that technical variation was adequately removed from the analysis. This approach is consistent with the normalisation strategy used in Chapter 4. This conservative normalisation approach increased the risk of losing small biological differences. It is possible that some more subtle differences in the GDM metabolome may have been eliminated by the normalisation strategy employed. If this was the case, it demonstrates a major issue in large metabolomics studies where the technical variation may be as high as the biological variation. However, this conservative approach offers added confidence in any significant differences that are observed. In this instance, the analyses did not reveal any significant differences. It is worth noting that some highly significant differences were observed in the plasma metabolome (Chapter 4), using the same conservative normalisation strategy. The hair metabolome, analysed using the methods detailed, appears to not be useful for differentiating GDM cases from controls.

Despite finding no significant metabolites in this study, there are some limitations that have been recognised, and some possibilities for ways that future investigations could be improved.

5.5 Conclusion

This study of the hair metabolome, in a large multi-ethnic Asian cohort, found no significant associations between the metabolites in hair, and the development of GDM. The biggest limitation of the study was the constraint in hair biomass that was collected at the time of bio-banking (2009/2010). Future studies of the hair metabolome would benefit from using standard guidelines for the collection of hair, such as those used in drug analysis of hair specimens (United Nations Office on Drugs and Crimes, 2014).

Chapter Six: Dietary Biomarkers of GDM

6.1 Introduction

Metabolomics can be used for understanding the association between maternal diet and GDM. The dietary associations with GDM reported thus far have predominantly been identified through observational research, capable of describing an association, but lacking the ability to explain the association. Metabolomics can be used to explain, from a metabolic perspective, *how* the diet is associated with GDM. The food metabolome is defined as the "part of the human metabolome directly derived from the digestion and biotransformation of foods and their constituents" (Scalbert et al., 2014, p. 1286). The food metabolome has been used to identify biomarkers of dietary intake, and to explore how different diets affect the metabolome. Much of the research into the food metabolome has been performed under controlled, clinical conditions (e.g., Garcia-Perez et al., 2017). Controlled conditions allow for a more reliable investigation and the determination of cause-and-effect relationships. However, these situations do not match the diversity of the diet consumed in a free-living population, and as such their results may not be applicable to the population to which they will be extrapolated. A smaller number of studies have investigated the food metabolome in an observational setting, relying on self-reported dietary data from FFQ, 24-hr recalls and other diet questionnaires to confirm associations (see Table 1.7). The data that has been generated in this field demonstrates an undeniable effect of diet on the metabolome, and combined with an abundance of research demonstrating an association between the maternal diet and GDM development (Appendix 1), there is huge potential to link dietary biomarkers with GDM. The concept behind these investigations is that if dietary biomarkers of GDM can be established, a prevention or management strategy might be easily unlocked in parallel. A biomarker of disease in its clinical application is only useful if there is a prevention, cure, or management strategy to employ. A biomarker for a disease without a known treatment may ultimately be useful for understanding the mechanism of disease, but in terms of the clinical use, this imposes a large cost of screening for no immediate health benefit. Once a dietary biomarker of disease is identified, a dietary intervention could be tailored and the biomarker tracked to explore whether the intervention could change the levels of these biomarkers, and therefore change the disease risk trajectory.

Only two studies to date have bridged the gap between dietary biomarkers and disease (detailed in Table 6.1).

Table 6.1

Studies of Dietary Biomarkers and Health Outcomes

Authors & date	Participants	Findings
Magnusdottir et al. (2014)	158 participants from the SYSDIET study, a randomised controlled multicentre trial of Nordic diet vs control diet (Kuopio and Oulu, Finland; Lund and Uppsala, Sweden; Aarhus, Denmark and Reykjavik, Iceland).	Total plasma alkylresorcinols (AR) concentration and C17:0/C21:0 homolog ratio was used as a biomarker of whole-grain rye intake (previously derived biomarker). The authors observed an inverse relationship between the AR ratio (biomarker of rye intake) in plasma and insulin concentration & the scores on insulin sensitivity indices (Matsuda ISI and DI). Results imply that a higher intake of whole-grain rye was associated with improved insulin sensitivity.
Mora-Cubillos et al. (2015)	Fifty participants from Spain with at least three Metabolic Syndrome risk factors that were part of a prospective, randomized, controlled, parallel-designed, 12-week mixed-nut intervention.	Using an untargeted LC-MS approach, the authors found Urolithin A glucuronide to be the most convincing biomarker of nut consumption in plasma. The authors found an inverse correlation between Urolithin A glucuronide and abdominal adiposity, and impaired glycemic control.

Both Magnusdottir et al. (2014) and Mora-Cubillos et al. (2015) used participants obtained from clinical trials where the intake of an individual food group or dietary pattern was monitored and controlled. Despite these studies demonstrating the ability to link dietary biomarkers to disease, both were essentially an a priori approach in terms of the dietary association – the diet was modified based on pre-established hypotheses. This type of analysis is used to understand how a dietary intervention relates to health outcomes via the metabolome, but is a reductionist approach. This approach would not necessarily be

appropriate for a free-living population, and cannot be achieved if the dietary association is elusive, as is the case with GDM. In GDM, the assessment of dietary biomarkers of disease guides hypothesis generation, it lets the data from free-living participants inform the dietary associations with the metabolome.

In previous chapters of this thesis, both the dietary information (in the form of derived dietary patterns), and the metabolome (hair and plasma) have been investigated for their associations with GDM in the GUSTO cohort. Dietary reporting is a subjective measure of dietary intake, whereas the metabolome is an objective measure of how the metabolism is affected in GDM. With evidence that diet significantly affects the metabolome (see Table 1.7) it is reasonable to hypothesise that dietary intake might significantly affect the plasma metabolome of individuals in this study. It was hypothesised that the association between the dietary patterns and GDM, found in Chapter 3, should be reflected in the GDM plasma metabolome of the individuals in this study (Chapter 4). Dietary assessment in the GUSTO cohort was established through a single 24-hr dietary recall which is a self-reported, observational measure of diet. The metabolome offers insight into *how/why* the diet is affecting the metabolism, and therefore how it might be related to the development of GDM. This may lead to a depth of understanding between maternal diet and GDM that has not been explored in previous studies.

The objective of this study was to investigate the intersect between the maternal diet and the plasma metabolome in regards to their associations with GDM, with the aim of identifying dietary biomarkers that could help to explain *how* the maternal diet is related to GDM.

6.2 Methods

6.2.1 Participants.

The participants in this study were a sub-set of participants from the GUSTO cohort. Details of the GUSTO study are reported in Chapter 2. GUSTO participants were eligible for this study if they met the following criteria:

a) Underwent an OGTT between the 26th and 28th weeks of gestation;

b) Completed a 24-hr dietary recall at the time of OGTT;

c) Had more than 200 µl of plasma available in the −80º C biobank;

d) Had complete data for maternal age, pregnancy BMI, and ethnicity (as these are three major clinical predictors of GDM in pregnancy and therefore need to be adjusted for in analyses).

6.2.2 Relationship between the seafood-noodle-based dietary pattern and plasma metabolites found to be associated with GDM.

The biomarkers of GDM that were found in the plasma metabolome in Chapter 4 are the starting point for these investigations. The 47 plasma metabolites that were significantly associated with GDM (after adjustment for confounding variables) were individually entered into a linear regression analysis with the seafood-noodle-based dietary pattern that was found, in Chapter 3, to be associated with GDM, to explore any associations between the dietary pattern and the plasma metabolites. The linear regression analyses were adjusted for the potential confounding variables (age, skinfolds, ethnicity, education, and previous GDM) and multiple comparison testing (Benjamini & Hochberg, 1995). Associations were considered statistically significant if the P – value was < 0.05 and FDR was < 0.05.

6.2.3 Canonical correlation of food groups and metabolites associated with GDM.

An alternative exploration of the association between maternal diet and GDM via the metabolome is to let the metabolomic data drive the investigation, rather than using the pre-formed ideas generated from the dietary pattern analysis results. A canonical correlation analysis was performed using the dietary dataset (68 food groups) and the metabolite dataset (47 plasma metabolites associated with GDM) to identify dietary biomarkers related to GDM. The roots produced by the canonical correlation represent the maximally correlated linear combinations of the two datasets. The canonical correlation roots were deemed significant if they had a P – value < 0.05.

6.3 Results

6.3.1 Participants.

A total of 787 participants met the eligibility criteria listed in Section 6.2. The participants who met the eligibility criteria for this study were the same participants reported in Chapter 4. The characteristics of the participants in this study are detailed in Table 4.3.

6.3.2 Relationship between the seafood-noodle-based dietary pattern and plasma metabolites found to be associated with GDM.

The seafood-noodle-based dietary pattern, which was found, in Chapter 3, to be associated with GDM, demonstrated a significant linear association with 11 of the plasma metabolites that were reported, in Chapter 4, to be associated with GDM (Table 6.2). After adjustment for confounding variables, two of the metabolites remained significantly associated with the seafood-noodle-based dietary pattern, both exhibiting a positive linear association (Table 6.3). Details of the metabolites, including their CAS number and library identification match factor can be found in Appendix 7. After adjustment for multiple comparisons, neither of the two metabolites remained statistically significant, with FDR values > 0.05.

Table 6.2

Metabolites Exhibiting a Significant Linear Association with the Seafood-Noodle-Based Dietary Pattern Prior to Adjustment

Metabolite	β	95% Confidence Interval	P - value
4-Methyl-2-oxopentanoic acid	0.07	(0.01, 0.13)	0.02
3-Methyl-2-oxopentanoic acid	0.07	(0.01, 0.13)	0.03
2-Hydroxybutyric acid	0.11	(0.04, 0.17)	< 0.001
Cyclobutene, 2-propenylidene-	0.11	(0.05, 0.18)	< 0.001
Norvaline derivative	0.13	(0.06, 0.20)	< 0.001
Heptadecanoic acid	0.17	(0.11, 0.24)	< 0.001
(E)-Tetradec-2-enal	0.17	(0.10, 0.23)	< 0.001
(7H)Thiopyrano[3,4-c]isoxazole, 3,3a,4,5-tetrahydro-7-(1-methylethyl)-	0.08	(0.01, 0.14)	0.02
Hydrazine, (3-methoxyphenyl)-	-0.09	(-0.16, -0.02)	0.01
Palmitoyl chloride	0.08	(0.01, 0.15)	0.02
Glycine derivative	0.17	(0.10, 0.24)	< 0.001

Table 6.3

Metabolites Exhibiting a Significant Linear Association with the Seafood-Noodle-Based Dietary Pattern After Adjustment for Confounding Variables

Metabolite	β	95% Confidence Interval	P - value	FDR
Heptadecanoic acid	0.11	(0.03, 0.18)	0.008	0.44
(E)-Tetradec-2-enal	0.17	(0.09, 0.17)	0.02	0.47

Regression analysis adjusted for: age, skinfolds, ethnicity, education, previous GDM

6.3.3 Canonical correlation of maternal diet (food groups generated from 24-hr recalls) and plasma metabolites associated with GDM.

The first two roots produced in the canonical correlation were statistically significant (root one $P < 0.001$; root two $P = 0.001$). The linear function of the food groups in the canonical correlation roots one and two could predict 37.1% and 31.0% of the variation in the linear function of the metabolites in their respective roots. However, the variance in the metabolites explained by the canonical correlation roots one and two was only 2% and 1.5%, respectively, and the variance in the food groups explained by the canonical correlation roots one and two was only 0.8% and 0.5%, respectively. The contents of the two roots and their correlation with the root are listed in Table 6.4 and Table 6.5. Details of the metabolites, including their CAS number and library identification match-factor can be found in Appendix 7. Box plots comparing the intake between GDM and non-GDM women of the food groups with a correlation to either of the significant canonical correlation roots > 0.3 or < -0.3 are displayed in Appendix 10 None of the three food groups contained in the canonical correlation roots were significantly associated with GDM ($P<0.05$).

Table 6.4

Metabolites and Food Groups in Root One Derived from the Canonical Correlation Analysis

Canonical correlation root one	Correlation
2-hydroxybutyric acid	-0.37
Norvaline	-0.45
Glycine	-0.53
Legumes/Pulses	0.42
Ethnic bread	0.56

Note: Only metabolites and food groups that had at least a moderate correlation with the canonical correlation root (> 0.3 or < - 0.3) were included in the table to represent root one.

Table 6.5

Metabolites and Food Groups in Root Two Derived from the Canonical Correlation Analysis

Canonical correlation root two	Correlation
White rice	-0.34
Arachidonic acid	-0.34
7,11-Hexadecadienal	-0.33

Note: Only metabolites and food groups that had at least a moderate correlation with the canonical correlation root (> 0.3 or < - 0.3) were included in the table to represent root two.

6.4 Discussion

This study was the first to investigate dietary biomarkers of GDM. The investigation was achieved by exploring the association between maternal diet (as determined by a 24-hr recall) and the maternal plasma metabolome of participants in the GUSTO study. It was hypothesised that the association between maternal diet and GDM, found in Chapter 3, would be reflected in the plasma metabolome via the metabolites found to be associated with GDM, in Chapter 4, as it has been shown previously that maternal diet influences metabolism and can alter the metabolomic profile.

Scores on the seafood-noodle-based dietary pattern (shown in Chapter 3 to be associated with a reduced likelihood of GDM development) demonstrated a positive linear association with the levels of two metabolites associated with GDM in Chapter 4 (heptadecanoic acid and (E)-tetradec-2-enal), after adjustment for confounding variables. After adjustment for multiple comparisons, both metabolite-dietary pattern associations were found to have a high FDR, indicating a high likelihood that both were false positive findings. Nevertheless, the positive association between heptadecanoic acid and the seafood-noodle-based dietary pattern is an interesting finding, given recent research. A study conducted on bottlenose dolphins found that a higher serum level of heptadecanoic acid was associated with lower serum ferritin levels, insulin, and triglyceride levels (Venn-Watson et al., 2015). That said, the dietary consumption of heptadecanoic acids in dolphins is predominantly through pinfish and mullet. Heptadecanoic acid is a long-chain saturated fatty acid that is exogenous in origin for humans, and is predominantly consumed through ruminant dairy products and meat (Wolk, Vessby, Ljung, & Barrefors, 1998), which have a very different overall nutrient profile to pinfish and mullet. The protective seafood-noodle-based dietary pattern that was associated with higher levels of heptadecanoic acid was not high in ruminant dairy, meat, or fish consumption, but was high in fish products such as fish balls and fish cakes which may have been a source of heptadecanoic acid. In Chapter 4 (E)-tetradec-2-enal was found to be in higher levels in women who developed GDM compared with those who did not develop GDM, whereas it was found to have a positive linear association with scores on the "protective" seafood-noodle-based dietary pattern. This relationship is not consistent with expectations and may be a false positive result due to multiple comparisons. However, (E)-tetradec-2-enal is a major volatile compound found in coriander (Shahwar et al., 2012), so it is highly plausible that coriander is consumed by women who scored highly on the seafood-

noodle-based dietary pattern, but also consumed by women who consumed other dietary patterns and developed GDM. It is unlikely that the consumption of coriander alone is responsible for the development of GDM, but that the combination of foods it is consumed with is important.

Based on the results of the linear regression of GDM-related plasma metabolites with the seafood-noodle-based dietary pattern, the plasma metabolome in this study did not appear to reflect the expected relationship between the maternal diet and GDM. There are some possible explanations for the discordant findings:

i) Firstly, the plasma metabolome was characterised at the time of OGTT, once the disease was already established, and therefore it is likely that the metabolome reflected the changes resulting from GDM rather than the changes leading to GDM development. Smaller influences on the metabolism, such as the diet, may have been overshadowed by the effect of established disease.

ii) Secondly, there are inherent limitations in using 24-hr dietary recalls to assess typical dietary intake. The best method of exploring the effect of the diet on the metabolome and its trajectory toward GDM development would be to track the maternal diet throughout the duration of pregnancy. Tracking the diet longitudinally would ensure that dietary biomarkers detected remain consistent throughout the different stages of pregnancy, and would also allow facilitate the earliest possible risk detection.

iii) Associations between diet and the metabolome related to GDM are complex, and an integrative approach to further investigations is necessary to understand other components that influence both the metabolome and the processing of nutrients. One avenue that could be useful to explore in conjunction with maternal dietary intake is the gut microbiome. The gut microbiome evolves rapidly in response to dietary exposure (David et al., 2013), and the microbiome differs between people with and without Type 2 diabetes, obesity, and in women with excessive gestational weight gain – a risk factor associated with GDM development (Hedderson et al., 2010; Larsen et al., 2010; Santacruz et al., 2010; Schwiertz et al., 2010). The gut microbiome may also play an important role in appetite regulation (Fetissov, 2016) and the availability and utilisation of energy from dietary intake (Krajmalnik-Brown, Ilhan, Kang, & DiBaise, 2012). Therefore, the gut microbiome may be involved in the association between diet and GDM.

The investigation of the intersect between the metabolites related to the habitual consumption of the protective seafood-noodle-based dietary pattern and the metabolites associated with GDM was essentially an a priori approach to try and understand the metabolic association between maternal diet and GDM. As an alternative to this method, a canonical correlation was performed to explore the relationship between the metabolites associated with GDM and the food groups constructed from the 24-hr recalls, to let the metabolomic data inform the associations. Two significant canonical correlation roots containing linear combinations of the dietary data (68 food groups) and the GDM-related plasma metabolome (47 metabolites) were constructed. These two roots, despite being statistically significant, explained only a very small portion of the variation in the food groups (1.3%) and the metabolite data (3.5%). The findings are not substantial enough to demonstrate that dietary biomarkers could be used to predict GDM development, but they do provide interesting information that could be explored in future studies to understand *how* the maternal diet influences metabolism, and how this could lead to GDM development. The canonical correlation root one consisted of three metabolites negatively correlated to the dimension (2-hydroxybutyric acid, norvaline, and glycine) and two food groups positively correlated to the dimension (legumes/pulses and ethnic bread). The metabolites in root one were found in higher levels in the participants who developed GDM (Chapter 4). The opposite direction of the correlation in the food groups and metabolites to the canonical correlation root implies that as scores of participants in root one increase, the amount of legumes/pulses and ethnic breads consumed increases, while the three metabolites decrease, and in theory, so would the likelihood of GDM. These findings are not consistent with the dietary pattern findings, where legumes/pulses and ethnic bread were negatively loaded on the protective seafood-noodle-based dietary pattern. However, these findings are in accord with previous research, where legumes/pulses, and grains were beneficial for glycaemic control and in reducing the risk of diabetes development (Venn & Mann, 2004; Yadav, Jain, & Sinha, 2007). Roti was one type of ethnic bread consumed in this cohort, a bread which is predominantly made from whole-wheat flour and therefore may be adding the benefits of fibre and lower glycaemic index when substituted for other carbohydrate staples such as white rice. Canonical correlation root two contained arachidonic acid, 7,11-hexadecadienal, and white rice, all of which were negatively correlated with the root. For each increase in a participant's score on root two, there was a reduced intake of white rice, as well as lower levels of arachidonic acid and 7,11-hexadecadienal. The direction of the relationship was the same for the two metabolites and white rice. White rice was one of the main food groups with a negative loading in the protective seafood-noodle-based dietary

pattern, supporting this association. These findings are particularly interesting as they link white rice intake in this cohort with increased levels of arachidonic acid and 7,11-hexadecadienal. Arachidonic acid (AA) is a pro-inflammatory omega-6 fatty acid (Kuehl & Egan, 1980), and a high AA/EPA (eicosapentaenoic acid) ratio has been found to be correlated with visceral fat in Japanese men (Inoue, Kishida, Hirata, Funahashi, & Shimomura, 2013). Due to the nature of this study (relative abundance of metabolites, not full quantitation), it is not appropriate to calculate ratios between metabolites, but the AA/EPA ratio is an interesting concept that might warrant investigation in future studies. The metabolite 7,11-hexadecadienal is a major volatile component found in the dried roots of Euphorbia kansui liou (*Euphorbiaceae*), a root commonly found in traditional Chinese medicines. Kansui, as it is often referred to, is used in Chinese medicines as a remedy for oedema, ascites, and asthma (J. Shen et al., 2016), but has also been found to cause inflammation (L. Zhang et al., 2009). Therefore, both arachidonic acid and 7,11-hexadecadienal may have a role in inflammation, indicating that the consumption of white rice may be related to GDM through its effect on inflammatory pathways. This is an interesting finding that could merit exploration in future studies concerned with the association between glycaemic index and GDM development – through its effect on inflammatory pathways.

6.5 Conclusion

This study was the first to investigate the relationship between maternal diet and GDM using a metabolomic approach. Using a canonical correlation analysis, some interesting associations were found between the maternal diet and the plasma metabolites related to GDM. Based on the findings, it could be hypothesised that legumes/pulses and ethnic breads consumed by this cohort were associated with a reduced risk of GDM through their negative relationship with 2-hydroxybutyric acid, norvaline, and glycine, three metabolites which were found in higher levels in the plasma of women who developed GDM. In contrast, white rice consumption could be hypothesised to be associated with an increased risk of GDM through its positive relationship with arachidonic acid and 7,11-hexadecadienal, two metabolites found to be higher in the plasma of women who developed GDM, both of which may have a role in increasing inflammation. Efforts to explain the association between the protective seafood-noodle-based dietary pattern and GDM using the plasma metabolome in this cohort were less successful. Two of the metabolites found to be higher in women who developed

GDM were found to have a significant positive linear association with scores on the seafood-noodle-based dietary pattern. However, these associations were no longer statistically significant after adjustments for multiple comparison testing, with FDRs > 0.4, implying the associations were likely to be false positive findings. Overall, despite the canonical correlation producing some interesting findings that could be explored in future studies, the strength of the associations was very low and therefore could not be deemed adequate to be used as dietary biomarkers of GDM development. It is plausible that the lack of strong, significant findings might be a result of the metabolome and diet being quantified at the time of OGTT, when disease was already established, making the interaction between diet and GDM difficult to distinguish from the metabolomic changes occurring as a result of GDM. Future work in this area should consider performing similar studies earlier in pregnancy, to investigate associations at a time where dietary changes could be made to prevent the development of GDM and its associated consequences. A longitudinal study would also be useful for investigating how the diet affects the maternal metabolome throughout the different stages of pregnancy, as it is not yet known whether dietary biomarkers are consistent throughout the course of pregnancy and when the earliest time would be to successfully intervene.

Chapter Seven: Summary and future directions

7.1 Summary of Findings

The aim of this thesis was to generate hypotheses on the relationship between maternal diet and GDM, by exploring the effect of both diet and disease on the metabolome. Dietary pattern analysis in conjunction with plasma metabolomic analysis, using participants from a large multi-ethnic Asian cohort – GUSTO (N=1247), resulted in some significant findings related to the associations between maternal dietary patterns and GDM development, and the metabolic dysfunction in GDM.

7.1.1 Maternal dietary patterns and GDM.

Firstly, as detailed in Chapter 3, exploratory factor analysis revealed three distinct dietary patterns consumed by participants in the GUSTO cohort: a vegetable-fruit-rice-based dietary pattern, a seafood-noodle-based dietary pattern, and a pasta-cheese-processed-meat-based dietary pattern. The seafood-noodle-based dietary pattern, consisting of higher amounts of soup, fish and seafood products, noodles in soup, flavoured noodles, red meat, seafood, and soya-sauce-based gravy; and lower amounts of white rice, curry-based gravy, legumes/pulses, and ethnic bread, was found to be significantly associated with a reduced likelihood of GDM development, after adjustment for confounding variables which included age, BMI, education, previous GDM, energy intake, and ethnicity ($P < 0.01$).

7.1.2 Plasma and hair metabolome in GDM.

Secondly, as detailed in Chapters 4 and 5, GC-MS analysis was conducted to characterise the plasma and hair metabolomes. Using the methods employed, a total of 219 metabolites were identified in the plasma metabolome, and 277 in the hair metabolome. Of the 219 plasma metabolites identified, 25 were found to be significantly higher in women who developed GDM after the adjustment for confounding variables and multiple comparisons, and one metabolite (9-octadecenamide) was significantly lower in women who developed GDM. A metabolic pathway analysis was conducted using online KEGG pathways, and a hypothesis was constructed linking a combination of the significant metabolites to a potential dysfunction in mitochondrial oxidative phosphorylation in the GUSTO participants who developed GDM. Some metabolites involved in inflammatory pathways and branched chain amino acid metabolism were also found to be associated with GDM development. The metabolite that was found to be most significantly different between GDM cases and non-

GDM controls, 2-hydroxybutyric acid (P-value = 7.04E-07), was higher in GDM cases and could discriminate GDM outcome with more accuracy than the model of clinical risk factors (age, BMI, ethnicity, family history of diabetes, previous GDM) typically used in practice to identify women at risk of GDM (ROC AUC= 0.71 vs 0.67). None of the 277 metabolites identified in the hair samples were found to be significantly different in women who developed GDM from those who did not, after adjustments for confounding variables and multiple comparisons.

7.1.3 Associations between maternal diet and the plasma metabolites related to GDM.

The significant plasma metabolomics findings were the key components used to explore the association between the maternal diet and GDM, by looking at how the maternal diet was related to the metabolic dysfunction observed in the plasma metabolome. In contrast to expectations that the "protective" seafood-noodle-based dietary pattern might be negatively related to the plasma metabolites identified in Chapter 4 as biomarkers of GDM, only two of the significant plasma metabolites demonstrated a linear association with the dietary pattern: heptadecanoic acid and (E)-tetradec-2-enal, both of which had a high FDR (> 0.4) and therefore, were likely to be a false positive result from multiple comparison testing. To further reduce assumptions and a priori expectations of associations between diet and disease, a canonical correlation was performed using the 68 food groups generated from the 24-hr dietary recalls and the plasma metabolites that were shown to be related to GDM, to explore whether any of the dietary components were linked to the dysregulated metabolic pathways observed in this set of participants. Two significant canonical correlation roots (containing metabolites and food groups) were constructed from the two datasets. Despite containing some interesting information that links the consumption of dietary components with metabolites found to be associated with GDM, the two canonical roots explained only 1.3% of the variation in the food groups and 3.5% of the variation in the metabolite data. Therefore, the findings are not strong enough to conclude the discovery of dietary biomarkers that could be used to predict GDM development, nor the identification of dietary components that have the potential to prevent GDM development.

7.2 Study Strengths and Impact of Findings

An undeniable strength of this collection of work was the ability to conduct investigations using a large, multi-ethnic, Asian cohort. A large cohort not only reduced bias and random effects, it also improved the representation of a population of interest, thereby improving the

translatability of findings, as well as improving the power to detect differences between groups. Very few studies conducted in relation to the maternal diet and the metabolome in GDM have used participants from Asian countries, despite Asian women having some of the highest rates of GDM worldwide. There are vast differences between Asian populations and Western populations (where a majority of the studies are conducted) in terms of genetics, culture, environment, diet, and metabolism; as such, it is difficult to justify the translation of Western findings to an Asian population. This was a gap in the literature that this study addressed. The significance of the difference between Western-based and Asian-based diets was evidenced in the dietary pattern findings. In the GUSTO dietary analysis, it was found that a seafood-noodle-based dietary pattern was protective against GDM development. This is a unique dietary pattern that had not been reported in any of the other studies investigating dietary patterns and GDM development. Interestingly, the vegetable-fruit-rice-based dietary pattern derived in the GUSTO cohort resembled a dietary pattern found in some of the previous literature (closely related to the "prudent" diet; Radesky et al., 2007; C. Zhang et al., 2006). However, unlike the previous studies, that pattern was not significantly associated with GDM in this multi-ethnic Asian cohort. These findings reinforce the importance of not translating dietary research findings beyond the scope of the population investigated.

An aspect often overlooked in previous studies of the metabolome in GDM was the adjustment for potential confounding variables. Without adjusting for confounding variables, particularly ones known to have an influence on metabolism, such as age and adiposity, it is likely that findings associated with GDM might be reflecting one or more of the underlying risk factors which are already well-known, adding no value to the field. As all of the confounding variables of interest in this study (age, adiposity [skinfolds and BMI], ethnicity, education, and previous GDM) were able to be identified or calculated using basic questionnaires or measurements, it would not be helpful to identify a metabolomic signature of any of these confounders. For example, if the metabolome in GDM without adjustment was simply reflecting age, this would be a very expensive and time-consuming process to identify a variable that could be extracted by simply asking the participant their age or inspecting their birth certificate. The work in this thesis placed importance on adjusting findings for confounding variables, strengthening the certainty of findings and their relationship with GDM.

The plasma metabolome findings in this work demonstrated the effect of GDM development on the plasma metabolome and, as such, support its use as a biofluid for exploring biomarkers

of GDM. The metabolite most significantly associated with GDM in this study, 2-hydroxybutyric acid, has a biologically plausible relationship with GDM development, has been found previously as an early marker of insulin resistance in a non-diabetic population, and was found in this study to be a better predictor of GDM diagnosis than the current clinical risk factors, emphasising its potential as a biomarker of GDM development (Gall et al., 2010). This finding should be validated in early pregnancy to assess its applicability as an early pregnancy biomarker to detect women at risk of GDM. Another strength of this study was that fasting plasma was analysed, which meant that significant findings were less likely to be attributed to recent dietary exposures, and any dietary associations observed with these metabolites would be likely to reflect habitual consumption.

The hair metabolome findings in Chapter 6 did not reproduce the findings of the previous work that was conducted on the hair metabolome in GDM (X. He et al., 2016). It did, however, highlight the importance of considering the study design for future work investigating the hair metabolome. The lack of significant findings, despite limitations in the study design, does pose the question of whether there are robust effects of GDM on the hair metabolome, as this study was sufficiently powered to observe differences and therefore it is possible that any differences were too subtle to detect. If this is the case, then the hair metabolome would not be deemed a suitable specimen for biomarker discovery and certainly not for use clinically, as a screening tool. The results from this study have led to some suggestions for improvements in study design for future studies of the hair metabolome (Chapter 5).

The novel component of this collection of work was the investigation of the metabolic mechanisms underlying the association between the maternal diet and GDM in free-living participants, using the plasma metabolome. It was found that the dietary pattern associated with GDM was not associated with the plasma metabolomic changes observed with GDM, as would have been anticipated. These findings could be a result of the dietary assessment methods employed: 24-hr dietary recalls may have been inadequate for characterising a "typical" diet and therefore its effect on overall metabolism; this is something that should be explored in future work. These unexpected findings did, however, lead to the investigation of a more objective approach to analysing dietary associations with GDM, using canonical correlation. The canonical correlation produced significant results that may shed some light on potential diet-metabolism associations related to GDM, but these findings need to be validated in a future study, using more robust dietary assessment measures.

7.3 Study Limitations

There are some limitations in the collection of work presented that are worthy of mention, and some suggestions for ways in which they could be addressed in future studies.

Firstly, the collection of work presented in this thesis was conducted using biological samples and data collected in the GUSTO study. Access to the GUSTO study was a great strength of this work as there are few studies conducted on Asian women, the cohort was from a population with a high-risk for GDM, and the cohort consisted of over 1000 women. It was also the only maternal cohort at the time of entering into this project that had collected biological specimens, demographic data, and dietary information from free-living pregnant women, as well as performing universal screening for GDM using an OGTT. Despite its strengths, the cohort design was not intended for this specific purpose and, as such, there were some limitations resulting from the data collection:

i) The dietary assessment measure chosen was a 24-hr dietary recall. For a large cohort study in which much data and specimens were collected from the participants, it is understandable that a 24-hr recall was chosen as it is far less time-consuming, for the recruitment staff as well as the participants, than the administration of a long, validated FFQ. However, 24-hr dietary recalls have been criticised for not capturing typical dietary diversity in free-living participants, and some studies have recommended that if a 24-hr dietary recall is used for assessment, at least four recalls on separate days are required to capture dietary variability, with one study recommending eight recalls (Holmes, Dick, & Nelson, 2008; Jackson, Byrne, Magarey, & Hills, 2008). The use of 24-hr recalls for dietary assessment, when compared with gold standards of dietary assessment which involve weighed-food diaries collected prospectively, is prone to recall bias, where participants can sometimes forget about food consumed in the previous 24 hr, particularly low volume additives such as spreads, and sweeteners added to hot drinks. Participants can also have difficulty accurately estimating quantities in retrospect. To combat these disadvantages, a small number of 3-day food diaries that were collected in a subset of women from the GUSTO study, were used to validate the dietary patterns derived from the 24-hr recalls. The validation did show a significant moderate correlation between scores on the dietary patterns generated using 24-hr recalls with applied scores constructed from the 3-day diaries, justifying the use of 24-hr recalls for dietary pattern analysis in

this study. It has also been shown previously that although a single 24-hr recall does not adequately represent habitual dietary micronutrient intake, percentage of macronutrient distribution is comparable to those calculated using 7-day weighed-food diaries (Rathnayake, Madushani, & Silva, 2012). However, to allow for the determination of accurate micronutrient and macronutrient associations in conjunction with dietary pattern associations to GDM, future studies should consider using an alternative method for dietary assessment. An FFQ, validated in the population of interest, would be a good compromise between food diary collection, which has very high subject burden but great accuracy, and 24-hr recalls, which have a much lower subject burden but are less representative of typical dietary consumption.

ii) Another limitation in using the GUSTO cohort was the use of biological samples and dietary data that were collected at the time of diagnosis. In an ideal situation, samples and information would have been collected longitudinally over the course of pregnancy, to determine how the metabolome changes over time in response to the progression to GDM, and the diet would be tracked in parallel. As this type of study is yet to be conducted in the population of interest, time of diagnosis was the next best option, prior to the participants being aware of their diagnosis. Maternal dietary patterns and diet quality have been found in previous work to not change significantly between pre-conception and during pregnancy, supporting the idea that the dietary pattern findings at time of diagnosis are likely to reflect dietary patterns consumed in early pregnancy (Cucó et al., 2006; Gresham, Collins, Mishra, Byles, & Hure, 2016; Crozier, Robinson, Godfrey, Cooper, & Inskip, 2009). The plasma metabolite that was found, in this work, to be significantly associated with GDM at time of diagnosis – 2-hydroxybutyric acid – has potential as an early pregnancy biomarker as it has been found to be an early marker of insulin resistance in a healthy non-pregnant population (Gall et al., 2010), supporting the use of resources to validate its application as a biomarker of GDM in early pregnancy. Despite samples and data collected at time of diagnosis not typically being used to explore biomarkers or preventative strategies, in the case of the work presented, this approach proved to be meaningful and provided some valuable insights for future work.

iii) The GUSTO study hair collection was not initially intended for metabolomics analysis and, as such, the collection protocols followed were not optimised for

metabolomic investigations. This resulted in a low hair biomass, no standardisation of hair length collected, and no standardisation of biomass across participants. Post-collection standardisation of length and biomass across participants was unable to be achieved due to the very small biomass that was available for analysis. The metabolite abundances were, however, normalised by biomass to statistically account for these differences. These limitations, combined with the conservative normalisation approach for batch correction, may explain why no metabolites were found to be significantly different between women who developed GDM and those who did not, compared to the previous study of the hair metabolome and GDM that found adipic acid to be significantly higher in GDM cases when compared to controls (X. He et al., 2016). In the study by X. He et al. (2016), the hair biomass was kept consistent across participants, and was much greater – at 10 mg per participant. It is possible that the lack of significant findings in this study was due to biomass limitations. However, as mentioned previously, it can be argued that if there were differences that were significant enough to be used in a clinical screening tool, they should have been robust enough to be found in a cohort of more than 700 women. Future studies should ensure that metabolomic researchers have input into the initial sample collection protocols prior to any large-scale pregnancy cohorts being established.

7.4 Future Directions

7.4.1 Short-term.

Based on the findings in this thesis, and the limitations observed, the next stage of the research would be to perform a validation study that addresses some of the limitations mentioned. The validation study should be conducted in early pregnancy to assess whether the dietary pattern findings are similar, and whether the potential plasma biomarker (2-hydroxybutyric acid) is able to predict women who will go on to develop GDM. A canonical correlation of dietary and metabolomics findings in early pregnancy, prior to diagnosis, may also uncover more interpretable associations because the metabolome will be less affected by disease and therefore, the effect of dietary intake might be easier to distinguish. A small validation study is planned using participants from the Healthy Mums and Babies (HUMBA) study (www.humba.ac.nz) - a randomised placebo-controlled intervention of probiotics and dietary advice during pregnancy to regulate blood sugar and reduce diabetes in pregnancy, in

an obese population in South Auckland, New Zealand. Although the study will be used to validate the biomarker findings from the GUSTO cohort, the dietary pattern findings will be a proof of concept whereby it would be expected, based on the GUSTO findings and the previous literature, that dietary patterns will differ from the unique multi-ethnic Asian population in GUSTO. The planned study will use the plasma samples collected from participants at recruitment (≤ 16 weeks' gestation) to perform both targeted and untargeted metabolomics, and dietary pattern analysis will be conducted using dietary intake assessed by a validated FFQ. Untargeted metabolomics of the plasma samples in the validation study will be performed following the same protocols as described in the thesis Chapter 4. In addition, a selection of metabolites (including 2-hydroxybutyric acid) will also undergo full quantitation to measure their plasma concentrations and determine their ability as early pregnancy biomarkers of GDM development. The recruitment of HUMBA participants is expected to be completed by the end of 2017.

7.4.2 Long-term.

The relationship between the maternal diet and GDM is extremely complex and to truly understand the effect of the diet on metabolism in relation to GDM development, a comprehensive longitudinal study should be conducted, provided the necessary resources and expertise can be obtained. The ideal future study would consist of a longitudinal prospective pregnancy cohort initiated in early pregnancy, with collection of data and samples at multiple time-points throughout pregnancy. Although it could be argued that pre-conception would be a better target to search for biomarkers and intervention strategies, this risks only providing such services to women who plan pregnancy. Research has demonstrated that up to 50% of pregnancies worldwide are unplanned (Sedgh, Singh, & Hussain, 2014), therefore early pregnancy – when women first confirm a pregnancy with their doctor or lead maternity carer, would be the best time to apply a screening tool that would reach the highest number of women who may require intervention. A longitudinal study would be used to investigate *how* and *when* the metabolome changes over the course of pregnancy, both in a healthy pregnancy and in the progression to GDM. This would assist with determining the earliest time point in pregnancy to detect women at risk of GDM development. It would also be used to define the metabolomic trajectory of a healthy (uncomplicated) pregnancy as it progresses throughout the trimesters – essentially like an infant-growth curve, but for healthy pregnancy progression. The healthy reference point or "curve" that is generated could then be used to assess how successful an intervention is to prevent GDM development, as the aim of an

intervention would be to reduce the risk by bringing an individual out of the "risk profile" and into the "healthy reference zone." To determine how the maternal diet could be manipulated to prevent GDM using data from the proposed longitudinal study, the following should be investigated:

1) What does the participant consume? This would entail measuring dietary intake using a validated FFQ and questionnaire on dietary supplement use. Ideally, the timing of intake would also be considered in some way, including how many meals are consumed in a day, and whether night-time feeding occurs.

2) What nutrients does the individual absorb and store from the diet reported, and at what levels? This would be more difficult to measure but could involve measuring their current nutrient profile in plasma or urine using comprehensive targeted nutrient panels. The reason for this component is that nutrient levels in a diet or supplement do not translate equally to plasma levels between individuals and therefore it is important to capture the effect of that diet on an individual, rather than hold the assumption that all participants consuming the same diet are absorbing the same levels of nutrients.

3) How does the microbiome interact with the diet? This would require genetic sequencing of the microbiome present in the faeces of participants as well as metabolomic profiling of the faecal samples to understand not just what microbes are present, but what metabolites they are producing in response to the diet consumed.

4) What is the effect of GDM on the metabolome? To comprehensively answer this question would require a metabolomics investigation using a range of biological specimens, collected at different time points throughout pregnancy, analysed using different extraction and derivatisation techniques, and different technologies, e.g., MS and NMR.

5) What is effect of the diet on the metabolome? This would consist of performing statistical analyses to relate the information gathered in 1-3 to the metabolomic information collected in 4, without consideration of GDM.

6) Most importantly, where does the effect of diet on the metabolome overlap with the trajectory towards GDM development? This involves identifying at which time point an overlap occurs and the details of how it overlaps. This is where the useful information for identifying a dietary intervention to prevent GDM could be found.

The findings from this proposed longitudinal study may be useful for generating population-based recommendations, after well-planned RCTs to show that the observational findings can be translated into beneficial effects following dietary modification. However, it is likely that the effect of diet on metabolism and its link to disease will be highly dependent on individual factors such as nutrient absorption rates, exercise, demographic characteristics, and the individual's gut microbiome, which could result in findings being more applicable for use in personalised nutrition.

Assuming success, these future studies would inform food-based interventions in pregnant women, aimed at preventing the development of GDM. The intervention would essentially be a personalised nutrition intervention where recommendations would be tailored based on an individuals' demographics (age, BMI, ethnicity), score for nutrient absorption, the make-up of their microbiome, and their current metabolomic state. Similar to the concept of a healthy growth chart for children, the intervention would be tailored to the individual to bring their metabolomic profile within the healthy set-points that have been determined from the longitudinal cohort study.

The complexity of the relationship between maternal diet and GDM means that more extensive studies are required to accurately determine how and why food and beverage consumption during pregnancy is associated with GDM development, and how the maternal diet can be manipulated to prevent GDM and its associated adverse consequences.

Appendices

Appendix 1. Studies of the maternal diet related to GDM between the years 2007-2017

Authors & Year	Dietary component	Participants	Assessment	Key findings
Grissaa et al. (2007)	Lipids, vitamins A, C, and E	Africa: 59 GDM cases 60 controls (age-matched)	Plasma/serum levels in third trimester	Higher blood levels of Vitamin C and lower levels of Vitamin E were observed in GDM cases compared to controls. Free cholesterol was significantly lower in GDM cases when compared with controls.
Hekmat et al. (2014)	Retinol	Iran: 41 GDM cases 41 controls	Serum levels in third trimester	Retinol lower in GDM cases.
Krishnaveni et al. (2009)	Vitamin B12	India: 49 GDM cases 736 Controls	Plasma B12 and folate from 30 weeks gestation	Deficient Vitamin B12 levels were associated with increased GDM rates, however this relationship was no longer significant after adjustment for maternal BMI.
Sukumar et al. (2016)	Vitamin B12 Folate	UK: 143 GDM cases 201 controls	Third trimester serum vitamin B12 and folate levels	Vitamin B12 levels were significantly lower in GDM cases compared to controls, after adjustment for confounding variables. There was no significant difference in folate levels between the two groups.
Bartáková et al. (2016)	Thiamine (Vitamin B1)	Czech Republic: 99 GDM cases 78 controls	Plasma levels	Vitamin B1 lower in GDM cases (adjusted for BMI).
Darling, Mitchell, and Werler (2016)	Iron	USA/Canada: 316 GDM cases	58-item food frequency questionnaire (FFQ)	Pre-conception dietary non-haem iron intake associated with lowered risk of GDM.

Authors & Year	Dietary component	Participants	Assessment	Key findings
		6913 controls		
Bowers et al. (2011)	Iron	USA: 876 GDM cases 12599 controls	133-item semi-quantitative FFQ	Pre-conception intake of haem iron associated with increased risk of GDM.
Bowers et al. (2016)	Iron	Denmark: 350 GDM cases 349 controls	Plasma ferritin and soluble transferrin receptor (sTfR) in early pregnancy (6-12 weeks' gestation)	After adjustment for confounders, including BMI, plasma ferritin remained significantly positively associated with GDM development. However, there was an interaction effect with smoking.
Khambalia et al. (2015)	Iron	Australia: 128 GDM cases 3648 controls	Serum ferritin and soluble transferrin receptor	The top tertile (≥ 35 μg/l) of ferritin level associated with an increased risk of GDM development after adjustment for confounding variables.
Rawal et al. (2017)	Iron	USA: 107 GDM cases 214 controls	Plasma hepcidin, ferritin and soluble transferrin receptor (sTfR) levels at gestational weeks 10–14 and 15–26	GDM cases had significantly higher ferritin levels at the first visit compared to matched controls. Hepcidin and ferritin were significantly higher in GDM cases compared to controls at the second visit, and sTfR:ferritin ratio was significantly lower in cases.
Omidvar et al. (2013)	Iron	Iran: 100 GDM cases 100 controls	Serum ferritin, TIBC, and serum iron levels at 26 weeks' gestation	Serum ferritin significantly higher in GDM cases compared to controls, and low ferritin level (< 20 ng/ml) significantly associated with a reduced risk of GDM.
Chan et al. (2009)	Iron	Hong Kong:	RCT:	No significant difference in prevalence of GDM

Authors & Year	Dietary component	Participants	Assessment	Key findings
		531 intervention 511 placebo	Intervention = 60mg iron / day	development between the two study groups.
Behboudi-Gandevani et al. (2013)	Iron Zinc	Iran: 72 GDM cases 961 controls	Serum zinc and iron levels Dietary intake assessed using semiquantitative FFQ Between 14 and 20 weeks of gestation	Serum iron levels were significantly higher in GDM cases when compared with controls.
Qiu et al. (2011)	Iron	USA: 158 GDM cases 3000 controls	FFQ to capture 3 months intake (pre-conception and first trimester)	Haem iron intake was significantly, positively associated with an increased risk of GDM.
Derbent et al. (2013)	Iron	Turkey: 30 GDM cases 72 controls	Serum hepcidin, iron, transferrin, transferrin saturation, and ferritin measured at 24–28 weeks of gestation.	Serum ferritin, hepcidin, and iron levels were significantly higher in GDM cases when compared to controls.
Helin et al. (2012)	Iron	Finland: 72 GDM cases 327 controls	Validated 181-item FFQ at 26-28 weeks' gestation	A borderline significant, positive association was observed between iron intake and GDM development, after adjustment for confounding variables.
Zein et al. (2015)	Iron	Lebanon: 16 GDM cases	Serum ferritin in first trimester	A significant positive association was observed between ferritin levels and 2-hr plasma glucose levels following OGTT, but no significant association was found with GDM

Authors & Year	Dietary component	Participants	Assessment	Key findings
		88 controls		as a categorical variable.
Akhlaghi, Bagheri, and Rajabi (2012)	Iron, Ni, Al, Cr, Mg, Zn, Cu, and Se	Iran: 30 GDM cases 30 controls	Plasma levels taken between 24 and 28 weeks' gestation	Iron levels significantly lower in GDM cases when compared to controls. No significant associations were found with any of the other micronutrients analysed.
Zhu et al. (2016)	Folic Acid	China: 249 GDM cases 1689 controls	Not specified	Folic acid supplementation in first trimester associated with an increased GDM risk
Shahgheibi, Farhadifar, and Pouya (2016)	Vitamin D	Iran (women with at least 1 risk factor for GDM): 46 intervention 46 controls	RCT 500 units of vitamin D/day OR placebo until 26 weeks' gestation	Vitamin D intervention significantly reduced the development of GDM when compared to the placebo-group (11.4% vs 34.8%)
Schneuer et al. (2014)	Vitamin D	Australia: 376 GDM cases 3714 controls	Serum 25-hydroxyvitamin D (25(OH)D) in first trimester	No association was observed between GDM and serum 25(OH)D in first trimester, after adjustment for confounding variables.
Zuhur et al. (2013)	Vitamin D	Turkey: 234 GDM cases 168 controls	Serum 25(OH)D at 24-28 weeks' gestation	Serum 25(OH)D was significantly lower in GDM cases when compared with controls. Severe deficiency (<12.5nmol/l) was associated with a significantly increased risk of GDM.
O. Wang et al. (2012)	Vitamin D	China: 200 GDM cases 200 controls	Serum 25(OH)D at 26-28 weeks' gestation	Serum 25(OH)D was significantly lower in GDM cases when compared with controls. Deficiency (< 25nmol/l) was associated with a significantly increased risk of GDM.

Authors & Year	Dietary component	Participants	Assessment	Key findings
McLeod et al. (2012)	Vitamin D	Australia: 63 GDM cases 336 controls	Serum 25(OH)D at 24-32 weeks' gestation	A significant negative association was observed between fasting plasma glucose levels and serum 25(OH)D levels, but there was no significant association with 1-hr or 2-hr post-OGTT plasma glucose levels.
Parlea et al. (2012)	Vitamin D	Canada: 116 GDM cases 219 controls	Serum 25(OH)D at 15-18 weeks' gestation	Serum 25(OH)D levels in early pregnancy were significantly lower in GDM cases when compared with controls. Serum 25(OH)D levels below the top quartile (< 73.5 nmol／l) were associated with an increased risk of GDM development, after adjustment for confounders of maternal age and weight.
Baker et al. (2012)	Vitamin D	USA: 60 GDM cases 120 controls	Serum 25(OH)D at 11 -14 weeks' gestation	No significant association between early pregnancy serum 25(OH)D levels and GDM were observed.
Savvidou et al. (2011)	Vitamin D	UK: 100 GDM cases 1000 controls	Serum 25(OH)D at 11 -14 weeks' gestation	No significant association between serum 25(OH)D at 11 - 14 weeks' gestation and GDM development.
Makgoba et al. (2011)	Vitamin D	UK: 90 GDM cases 158 controls	Serum 25(OH)D in first trimester	No significant association between early pregnancy serum 25(OH)D levels and GDM were observed.
C. Zhang et al. (2008)	Vitamin D	USA: 57 GDM cases 114 controls	Plasma 25(OH)D at 16 weeks' gestation	Plasma 25(OH)D concentrations were significantly lower in GDM cases than controls, even after adjustment for confounders. There was a significant negative linear association between plasma 25(OH)D levels and GDM risk (each 5

Authors & Year	Dietary component	Participants	Assessment	Key findings
				ng/ml decrease in 25(OH)D was associated with a 1.36-fold increase in GDM risk).
Farrant et al. (2009)	Vitamin D	India: 34 GDM cases 559 controls	Serum 25(OH)D at 30 weeks' gestation (± 2 weeks)	No significant association found between serum 25(OH)D levels and GDM.
Rodriguez et al. (2015)	Vitamin D	Spain: 93 GDM cases 2265 controls	Plasma 25-hydroxyvitamin D3 in early pregnancy	No significant association found between plasma 25-hydroxyvitamin D3 in early pregnancy and GDM development.
Kramer et al. (2014)	Vitamin D	Canada: 142 GDM cases 382 non-GDM controls	Serum 25(OH)D in late second trimester	No significant association between maternal vitamin D status and GDM.
Zhou et al. (2014)	Vitamin D	China: 331 GDM cases 1622 controls	Serum 25(OH)D at 16th-20th weeks gestation	Lower levels of serum 25(OH)D were associated with an increased risk of GDM development.
Domaracki et al. (2016)	Vitamin D	Poland: 103 GDM cases 36 controls	Serum 25(OH)D	No significant difference in serum 25(OH)D levels between GDM cases and controls.
Boyle et al. (2016)	Vitamin D	New Zealand: 32 GDM cases 1512 controls	Serum 25(OH)D at 15 weeks' gestation	Serum 25(OH)D levels were not significantly associated with GDM after adjustment for confounders.
Haidari et al. (2016)	Vitamin D	Iran: 45 GDM cases	Serum 25(OH)D collected between 20 and 30th	Serum 25(OH)D was significantly lower in GDM cases after adjustment for confounders.

Authors & Year	Dietary component	Participants	Assessment	Key findings
		45 controls	weeks' gestation	
Baker et al. (2012)	Vitamin D	USA: 60 GDM cases 120 controls	Serum 25(OH)D	Vitamin D status was not associated with GDM
Yap et al. (2014)	Vitamin D	Australia: 89 high dose vitamin D (5000 IU) 90 low dose (400 IU) vitamin D	RCT high or low dose vitamin D supplement daily starting in early gestation	High dose vitamin D supplementation did not improve glucose levels in pregnancy.
Dodds et al. (2016)	Vitamin D	Canada: 395 GDM cases 1925 controls	Serum 25(OH)D prior to 20 weeks' gestation	Lower levels of serum 25(OH)D in early pregnancy were associated with an increased risk of GDM.
Lacroix et al. (2014)	Vitamin D	Québec: 54 GDM cases 655 controls	Blood levels of 25OHD2 and 25OHD3	Low levels of 25OHD in first trimester associated with higher risk of GDM development
Loy et al. (2015)	Vitamin D	Singapore: 155 GDM cases 785 controls	Serum 25(OH)D at 26th-28th weeks' gestation	Lower levels of serum 25(OH)D were associated with a higher fasting plasma glucose level.
Mojibian et al. (2015)	Vitamin D	Iran: 246 intervention A 224 intervention B At gestational age 12-16	RCT Intervention A = 400 IU vitamin D (Cholecalciferol) daily	The high-dose intervention group had a significantly lower rate of GDM than the other intervention (6.7% vs 13.4%).

Authors & Year	Dietary component	Participants	Assessment	Key findings
		weeks' gestation with serum 25 (OH) D less than 30 ng/ml.	Intervention B = 50,000 IU every 2 weeks orally	
Arnold et al. (2015)	Vitamin D	USA AND Europe: 135 GDM cases 573 controls	Serum 25(OH)D and 25[OH]D3 in early pregnancy	No significant differences between cases and controls after adjustment for confounders.
Pleskačová et al. (2015)	Vitamin D	Czech Republic: 47 GDM cases 29 controls	Plasma 25(OH)D taken at 24th-30th weeks' gestation	No significant difference between plasma 25(OH)D levels of cases and controls.
S. Park et al. (2014)	Vitamin D	Korea: 23 GDM cases 500 controls	Serum 25(OH)D at 12th-14th weeks' gestation	No significant difference in serum 25(OH)D levels was observed between cases and controls.
Flood-Nichols et al. (2015)	Vitamin D	USA: 5 GDM cases 235 controls	Plasma 25(OH)D in first trimester	Vitamin d levels were not significantly associated with GDM development.
Nobles, Markenson, and Chasan-Taber (2015)	Vitamin D	USA: 14 GDM cases 135 controls	Serum total 25(OH)D levels	Higher 25(OH)D was associated with an increased risk of GDM in Hispanic women
Whitelaw et al. (2014)	Vitamin D Calcium	UK: 137 GDM cases 1330 controls	Serum calcium, 25(OH)D2, and 25-dihydroxyvitamin D3 at 25-27 weeks' gestation	No significant association found between vitamin D and GDM. Calcium levels were positively associated with GDM diagnosis, and this remained significant after adjustment for confounders.

Authors & Year	Dietary component	Participants	Assessment	Key findings
Osorio-Yáñez et al. (2016)	Calcium	USA: 169 GDM cases 3245 controls	Validated semi-quantitative FFQ at 15 weeks to assess preconception and first trimester diet	Calcium intake was not significantly different between cases and controls. There was a significant trend between the increased consumption of both wholegrains and low-fat dairy (analysed independently) and reduced risk of GDM development.
Goker Tasdemir et al. (2015)	Magnesium	Turkey: 40 GDM cases 45 controls	Magnesium levels at 26-28 weeks' gestation	Magnesium levels were significantly lower in GDM cases when compared to controls.
Farland, Rifas-Shiman, and Gillman (2015)	Food cravings	USA: 117 GDM cases 1905 controls	Validated FFQ in first trimester to measure intake since last menstrual period	Salty food cravings were associated with a significantly reduced risk of GDM.
Ley et al. (2013)	Vitamin E	USA: 47 GDM cases 148 controls	FFQ delivered at 30 weeks' gestation to retrospectively assess second trimester intake	Lower dietary intake of vitamin E was associated with a higher fasting glucose level and measure of insulin resistance.
Hamdan et al. (2014)	Zinc Selenium	Sudan: 31 GDM cases 31 controls	Serum zinc and selenium levels (at approx.. 33 weeks' gestation)	No significant association between zinc or selenium levels and GDM.
Kilinc et al. (2008)	Selenium	Turkey: 30 GDM cases 101 controls	Serum selenium levels between 24 and 28 weeks of gestation	Serum selenium levels were significantly lower in GDM cases when compared to controls.

Authors & Year	Dietary component	Participants	Assessment	Key findings
Bowers et al. (2012)	Fat	USA: 860 GDM cases 12615 controls	133-item semi-quantitative FFQ	Higher intake of cholesterol, animal fat, and mono-unsaturated fatty acids in preconception were associated with increased risk of GDM.
X. Chen et al. (2010)	Fat	USA: 49 GDM cases 98 controls	24-hr recalls at 15, 20, and 28 weeks' gestation Serum fatty acid levels of myristic, palmitic, stearic, palmitoleic, oleic, linoleic, linolenic, arachidonic, eicosapentenoic (EPA), and docosahexenoic (DHA)	15 week and third trimester levels of stearic acid were significantly lower in GDM cases when compared to controls. 15 week levels of palmitoleic acid were significantly higher in GDM cases when compared to controls. Third trimester total saturated fatty acids were significantly lower in GDM cases, whereas oleic acid and total mono-unsaturated fatty acids (MUFA) were significantly higher. The only significant correlation between dietary intake and serum levels was for polyunsaturated fatty acids (PUFA). Dietary intake of PUFA, linoleic acid, and DHA were significantly lower in GDM cases when compared to controls, whereas the dietary intakes of total saturated fatty acids, palmitic acid, and stearic acid were significantly higher in GDM cases.
Barbieiri et al. (2016)	Fat	Brazil: 151 GDM cases 648 controls	Two non-consecutive day 24-hr dietary recalls	Higher levels on the thrombogenicity indices (TI) were associated with an increased likelihood of GDM, and higher hH indices (ratios of hypo-and hypercholesterolemic fats) were associated with a lower likelihood of GDM

Authors & Year	Dietary component	Participants	Assessment	Key findings
S. Park et al. (2013)	Fat Energy Intake Taurine	Korea: 215 GDM cases 531 controls	24 hr recall at 24–28th week of pregnancy	Energy intake was significantly higher in GDM cases when compared to controls. Fat intake as a % of energy was significantly lower in GDM cases when compared to controls (but authors did not adjust for energy intake). Saturated fatty acid intake was significantly associated with GDM development using a backwards stepwise logistic regression, in normal weight women. However, this was not adjusted for energy intake or other confounding variables. Taurine intake was significantly lower in GDM cases compared to controls, although this was not adjusted for energy intake differences between groups.
D'Anna et al. (2015)	Myo-inositol	Italy (BMI > 30): 110 treatment group 110 placebo	RCT (2g twice a day) from first trimester until delivery	Women in the treatment group were significantly less likely to develop GDM (14% vs 34%)
D'Anna et al. (2012)	Myo-inositol	Italy (all women had Poly cystic ovary syndrome (PCOS)): 46 treatment group 37 control group	Clinical trial Treatment = 4 g per day, beginning pre-pregnancy until delivery Control = metformin pre-pregnancy, stopped in pregnancy	Myo-inositol supplementation significantly reduced the number of women with PCOS that developed GDM, when compared to controls (17% vs 54%).
D'Anna et al. (2013)	Myo-inositol	Italy (all women had parent with history of type 2 diabetes):	RCT (2g twice a day) from end of first trimester until delivery	Women in the treatment group were significantly less likely to develop GDM (6% vs. 15%)

Authors & Year	Dietary component	Participants	Assessment	Key findings
		110 treatment group 110 placebo		
Murphy et al. (2016)	Myo-inositol	USA: 35 GDM cases 59 controls	Urinary levels at 6 - 14 weeks' gestation and 22 - 32 weeks' gestation	Early pregnancy urinary levels of myo-inositol were significantly higher in women who went on to develop GDM compared to controls. No significant difference was observed in later pregnancy.
Santamaria et al. (2016)	Myo-inositol	Italy: 30 GDM cases 30 controls	Amniotic fluid levels between 15th and 18th weeks of gestation	Women who later developed GDM, had higher levels of myo-inositol in the amniotic fluid at 15-18weeks gestation than the controls.
Ley et al. (2011)	Macronutrients	Canada: 47 GDM cases 158 controls	Validated FFQ to recall dietary intake in second trimester	Total fat, saturated fat, trans fat (as % of energy intake), and added sugar in coffee and/or tea were all significantly positively associated with fasting plasma glucose levels, after adjustment for confounders. Vegetable and fruit fibre was significantly inversely associated with fasting plasma glucose levels, after adjustment for confounders. Increased intake of carbohydrate (as % of energy) was associated with a reduction in GDM risk per 1-SD increase, and total fat (as % of energy) was associated with an increased risk of GDM per 1-SD increase, after adjustment for confounders.
Qiu et al. (2011)	Eggs Cholesterol	USA: 158 GDM cases 3000 controls	121-item semi-quantitative food frequency questionnaire	> 10 eggs/week increased risk of GDM development when compared to no egg consumption. Egg consumption > 7/week increased risk of GDM 1.8 fold when compared with egg consumption < 7/week

Authors & Year	Dietary component	Participants	Assessment	Key findings
		238 GDM cases 502 controls		The highest quartile of cholesterol intake was associated with increased risk of GDM (>294) > 7 eggs/week increased risk of GDM development when compared to no egg consumption. Egg consumption > 7/week increased risk of GDM 2.7 fold when compared with egg consumption < 7/week The highest quartile of cholesterol intake was associated with increased risk of GDM (> 294)
Bao et al. (2016)	Potatoes	USA: 854 GDM cases 20839 controls	Questionnaire	Total consumption of potatoes pre-pregnancy demonstrated a positive, significant association with an increased risk of GDM
Bao et al. (2014)	Fried foods	USA: 847 GDM cases 20232 controls	Validated FFQ pre-pregnancy	High fried food consumption (> 7/week) was associated with an increased risk of GDM development, after adjustment for confounders. Fried food consumption at home was not significantly associated with GDM development, but fried food eaten away from home was.
Sahariah et al. (2016)	Leafy Green Vegetables, Fruit, and Milk	India: 100 GDM, 908 non-GDM 492 Intervention group 516 control group	RCT pre-conception and throughout pregnancy: Intervention group = Daily Snack Containing Leafy Green Vegetables, Fruit, and Milk Control group = low-micronutrient vegetables	Significantly less women in the treatment group developed GDM when compared to the control group (7% vs 12%). The significance remained after adjustments for BMI and weight gain during pregnancy.

Authors & Year	Dietary component	Participants	Assessment	Key findings
			(potatoes, onions)	
Salmenhaara et al. (2010)	Food groups, and nutrient consumption	Finland: 174 GDM cases 3613 controls	Semi-quantitative, validated food frequency questionnaire in Eighth month pregnancy	Consumption of milk products, cereal products, vegetables, and meat were significantly higher in cases when compared to controls. Protein and dietary fibre was higher in cases, sugars and saturated fat were significantly lower in cases. Vitamin A, D, folate, and iron were significantly higher in cases also.
Bao et al. (2013)	Protein	USA: 870 GDM cases 20587 controls	Semiquantitative FFQ pre-pregnancy	Animal protein intake was associated with an increased risk of GDM after adjustment for confounding variables, whereas vegetable protein was associated with a reduced risk of GDM development after adjustment for confounding variables. Total red meat and unprocessed red meat were two major protein food groups that were significantly associated with an increased risk of GDM, whereas nuts were associated with a reduced risk.
Karamanos et al. (2014)	Dietary Pattern	10 different Mediterranean countries (Algeria, France, Greece, Italy, Lebanon, Malta, Morocco, Serbia, Syria and Tunisia): 95 GDM cases 908 controls	78-question FFQ	An association was found between increased adherence to a Mediterranean dietary pattern and reduced likelihood of GDM development
Izadi et al. (2016)	Dietary	Iran:	Three 24-hr dietary	GDM cases had significantly lower adherence to the DASH

Authors & Year	Dietary component	Participants	Assessment	Key findings
	approaches to stop hypertension (DASH) diet Mediterranean diet	200 GDM cases 263 controls	recalls used to calculate scores for DASH and MED diet adherence (participants recruited between 5 and 28 weeks' gestation)	diet when compared with controls, whereas there was no significant difference in adherence to the MED diet.
Bao et al. (2014)	Low-carbohydrate diet	USA: 867 GDM cases 20544 controls	Pre-pregnancy Low Carbohydrate Diet (LCD) scores calculated from validated food-frequency questionnaires	A pre-pregnancy LCD with protein and fat sources predominantly from animal foods showed a positive significant association with GDM risk.
Markovic et al. (2016)	Glycaemic Index	Australia: (At least one risk factor for GDM: age >35 years, first-degree relative with T2DM, prepregnancy BMI ≥30 kg/m2, past history of GDM or glucose intolerance, history of a previous baby >4,000 g, or belonging to a high-risk ethnic group (Aboriginal or Torres Strait Islander, Polynesian, Middle Eastern, Indian, or Asian))	RCT (recruited between 12 and 20 weeks of gestation) Intervention = Low-GI diet Control = High-fibre, moderate-GI diet	The Low-GI dietary intervention did not reduce the risk of GDM in this high-risk population.

Authors & Year	Dietary component	Participants	Assessment	Key findings
		65 intervention group 60 control group		
Li et al. (2015)	Dietary patterns	China: 74 GDM cases 497 controls	Validated FFQ to assess diet between second trimester and time of OGTT (24-28th week of gestation)	Women with GDM had a significantly higher intake of high-GI fruits and energy-dense snacks and were associated with increased likelihood of GDM in regression analyses.
Li et al. (2014)	Dietary behaviours	China: 69 GDM cases 470 controls	81-item FFQ to evaluate dietary intake from beginning of the second trimester of pregnancy (12th gestational week) to the time of OGTT (24th to 28th gestational week)	There were no significant differences in macronutrient composition of the diet between the case and control groups. GDM cases consumed a significantly higher amount of high GI fruits and energy dense snack foods.
S. Park et al. (2013)	Dietary behaviours Nutrient analysis	Korea: 44 GDM cases 219 controls	Dietary behaviours questionnaire, 3-day food record for nutrient consumption (2 week days and 1 weekend day)	Coffee intake was lower in GDM cases, but black coffee intake was higher in controls, whereas GDM cases tended to consume a mixed variety of coffee and/or coffee with additional sugar. Night-time eating was higher in GDM cases when compared to controls. Energy intake in GDM cases was significantly higher than controls. Energy-adjusted intake of calcium, sodium, and vitamin B2 were significantly higher in controls, whereas energy-

Authors & Year	Dietary component	Participants	Assessment	Key findings
				adjusted intakes of vitamin C and SFA were significantly higher in GDM cases.
Loy et al. (2017)	Timing of food consumption	Singapore: 198 GDM cases 863 controls	Interviews at 26-28 weeks' gestation + 24-hr dietary recall	No association between night-fasting hours or eating episodes per day and GDM. However, eating episode was positively, significantly associated with 2hr glucose measures after OGTT, after adjustment for confounders.
Meinila et al. (2017)	Diet quality	Finland: 29 GDM cases 108 controls	First trimester FFQ to derive score on healthy food intake index to reflect adherence to Nordic nutrition recommendations	Score on healthy food intake index was not significantly associated with GDM risk.
Gresham et al. (2016)	Dietary Quality	Australia: 83 GDM cases 1824 controls	Validated 74-item FFQ (to determine previous 12-months intake). Collected in pre-pregnancy for 38% of participants and during pregnancy for the remaining 62%	Score on the Australian recommended food score was not significantly associated with GDM outcome.
Bartáková et al. (2016)	Food groups	Czech Republic: 293 GDM cases 70 controls	Semi-quantitative FFQ to determine diet intake 6 months before OGTT (conducted between 24th-30th weeks' gestation)	High-protein foods and sugar-sweetened beverages, dairy products, smoked meat, pork, poultry, fresh fruit, syrup, and confectionary were significantly higher in GDM cases compared to controls. Fresh vegetable consumption was significantly lower in GDM cases. No adjustments were made for confounders eg. BMI

Appendix 2. The Advantages and Limitations of NMR and MS in Metabolomics

Analytical Platform	Strengths	Limitations
Nuclear Magnetic Resonance	• Non-destructive • Highly reproducible • Requires less sample preparation • High throughput • Quantitative	• Expensive by comparison • Large – require significant amount of space • Limited sensitivity – more sample required • Destructive • Laborious sample pre-processing before analysis e.g., extraction and derivatisation
Mass spectrometry	• High specificity • High sensitivity • Small sample volume ok • Smaller by comparison – benchtop equipment • Can detect 100s-1000s of metabolites when combined with different chromatographic separation techniques and sample extraction and derivatisation	

References: (Dunn et al., 2011; Emwas, 2015; Pan & Raftery, 2007)

Appendix 3. Plasma collection SOP

GUSTO Protocol for Processing Maternal Blood

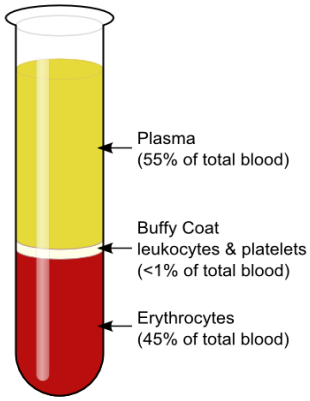
(version 1 September 2009)

Note: All samples should be kept at 4°C, for minimum amount of time. Handle all samples on ice at all times. Work as fast as possible. Do not leave samples at room temperature.

- A total of 20 ml blood is collected in 2 x 10ml EDTA tubes (labeled tubes 1 and 2). Process all blood collected within 4 hours.

- Immediately, mix thoroughly the blood collected in tube 1. Transfer 0.4 ml of whole blood into a clean 2ml cryovial tube. Add 1.2 ml Trizol, mix thoroughly and store at −80°C. Process the remaining blood in tube 1 together with tube 2.

- Centrifuge blood at 4°C, 3000 rpm for 10 min. There should be 3 distinct separate layers: plasma (upper layer), buffy coat (middle layer), and red cells (lower portion).



- Carefully retrieve plasma layer, taking care not to take too much so as not to disturb the buffy coat layer (may be a very thin layer). Transfer crude plasma into 1.5ml eppendorf tubes, and centrifuge at 4°C, 13200 rpm (Eppendorf® 5415R) for 10 min.

A. Aliquoting of Plasma

- Approximately 8 ml of plasma will be retrieved; divide into aliquots as follows (keep in 4°C at all times):

1. *For hormone analysis*
   - 4 ml of plasma is added to plain tube with trasylol
   - **Aliquots: 10 x 0.4ml of plasma**

2. *For RNA analysis*
   - **Aliquots: 4 x 0.4 ml of plasma**
   - Add 1ml of trizol to each tube

3. *For DNA analysis*
   - **Aliquots: 6 x 0.4ml of plasma without additive**

B. Aliquoting of Buffy Coat

- Retrieve the buffy coat portion (can have a bit of carry over from plasma layer above and red cell layer below) from each EDTA tube. Divide into aliquots as follows (keep in 4°C at all times):
- **Aliquots: 1 x 0.4ml of buffy coat**

C. Aliquoting of Red Cells

- Divide the red cell portion (~6ml) into aliquots as follows (keep in 4°C at all times):
- **Aliquots: 4 x 2ml of red cells**

Store all aliquots in −80℃ immediately.

Appendix 4. Plasma metabolomic extraction SpeedVac drying times

Sample Number	First speedvac time	Second speedvac time	Total speedvac time
1-18	5 hours, 20 mins	2 hours	7 hours, 20 mins
19-72	5 hours	2 hours	7 hours
73-162	3 hours, 40 mins	3 hours, 20 mins	7 hours
163-284	5 hours, 30 mins	1 hour, 30 mins	7 hours
285	5 hours, 30 mins	5 hours, 30 mins	11 hours
286-290	5 hours, 30 mins	1 hour, 30 mins	7 hours
291	5 hours, 30 mins	5 hours, 30 mins	11 hours
292-301	5 hours, 30 mins	1 hour, 30 mins	7 hours
302	5 hours, 30 mins	5 hours, 30 mins	11 hours
303-304	5 hours, 30 mins	1 hour, 30 mins	7 hours
305	5 hours, 30 mins	5 hours, 30 mins	11 hours
306	5 hours, 30 mins	1 hour, 30 mins	7 hours
307-310	7 hours, 33 mins	2 hours, 30 mins	10 hours, 3 mins
311	4 hours	3 hours	7 hours
312	7 hours, 33 mins	2 hours, 30 mins	10 hours, 3 mins
313	4 hours	3 hours	7 hours
314-322	7 hours, 33 mins	2 hours, 30 mins	10 hours, 3 mins
323-324	4 hours	3 hours	7 hours
325	7 hours, 33 mins	2 hours, 30 mins	10 hours, 3 mins

Sample Number	First speedvac time	Second speedvac time	Total speedvac time
326	4 hours	3 hours	7 hours
327	7 hours, 33 mins	2 hours, 30 mins	10 hours, 3 mins
328	4 hours	3 hours	7 hours
329-330	7 hours, 33 mins	2 hours, 30 mins	10 hours, 3 mins
331-332	4 hours	3 hours	7 hours
333-339	7 hours, 33 mins	2 hours, 30 mins	10 hours, 3 mins
340-380	4 hours	3 hours	7 hours
381	7 hours, 33 mins	2 hours, 30 mins	10 hours, 3 mins
382-435	4 hours	3 hours	7 hours
436	4 hours	4 hours	8 hours
437-680	4 hours	3 hours	7 hours
681	4 hours	4 hours	8 hours
682-711	4 hours	3 hours	7 hours
712	4 hours	4 hours	8 hours
713-741	4 hours	3 hours	7 hours
742	4 hours	4 hours	8 hours
743-750	4 hours	3 hours	7 hours
751	4 hours	4 hours	8 hours
752-985	4 hours	3 hours	7 hours
986	4 hours	4 hours 40 mins	8 hours, 40 mins

Sample Number	First speedvac time	Second speedvac time	Total speedvac time
987	4 hours	3 hours	7 hours
988	4 hours	4 hours 40 mins	8 hours, 40 mins
989-991	4 hours	3 hours	7 hours
992	4 hours	4 hours 40 mins	8 hours, 40 mins
993-998	4 hours	3 hours	7 hours
999	4 hours	4 hours 40 mins	8 hours, 40 mins
1000-1001	4 hours	3 hours	7 hours
1002-1003	4 hours	4 hours 40 mins	8 hours, 40 mins
1004-1005	4 hours	3 hours	7 hours
1006	4 hours	4 hours 40 mins	8 hours, 40 mins
1007-1013	4 hours	3 hours	7 hours
1014	4 hours	4 hours 40 mins	8 hours, 40 mins
1015-1022	4 hours	3 hours	7 hours
1023-1024	4 hours	4 hours 40 mins	8 hours, 40 mins
1025	4 hours	3 hours	7 hours
1026	4 hours	4 hours 40 mins	8 hours, 40 mins
1027-1029	4 hours	3 hours	7 hours
1030	4 hours	4 hours 40 mins	8 hours, 40 mins
1031-1035	4 hours	3 hours	7 hours
1036	4 hours	4 hours 40 mins	8 hours, 40 mins

Sample Number	First speedvac time	Second speedvac time	Total speedvac time
1037-1044	4 hours	3 hours	7 hours
1045	4 hours	4 hours 40 mins	8 hours, 40 mins
1046-1086	4 hours	3 hours	7 hours

Appendix 5. Reference compounds in in-house library used for identification in phase one of the plasma and hair metabolite identification steps

10,12-octadecadienoic acid (C18_2n-10,12c)	bishomo-gamma-Linolenic acid (C20_3n-6,9,12c)	Linoleic acid (C18_2n-6,9c)
10,13-dimethyltetradecanoic acid (C17_0)	c13-Glycine	Linolelaidic acid (C18_2n-9,12c)
10-Heptadecenoic acid (C17_1n-7t)	Cabamic acid	Lysine
10-Pentadecenoic acid (C15_1n-5c)	Caffeine	Malic acid peak 1
11,14,17-Eicosatrienoic acid (C20_3n-3,6,9c)	Caprinoic acid	Malic acid peak 2
11,14-Eicosadienoic (C20_2n-6,9c)	Caprylic acid	Malonic acid
13,16-Docosadienoic acid (C22_2n-6,9c)	cis-4-Hydroxyproline	Margaric acid (C17_0)
14.9814 min Dimethyl aminomalonic acid	cis-Aconitic acid	Methionine
16.2858 min (-)-O-Acetylmalic anhydride	cis-Vaccenic acid (C18_1n-7c)	Methoxytryptophan
1-Aminocyclopropane-1-carboxylic acid	Citraconic acid	Methylthioacetic acid
1-Phenylethanol	Citramalic acid	Myo-inositol Acetylated
2,3-Butanediol	Citric acid	Myristic acid (C14_0)
2,4-Diaminobutyric acid	Citric acid secondary peak	Myristoleic acid (C14_1n-5c)
2,4-Di-tert-butylphenol (derivatization artefact)	Conjugated linoleic acid (C18_2n-9,11c)	N-Acetylcysteine
2,6-Diaminopimelic acid	Coumaric acid	N-Acetylglutamic acid
21.4162 min L-(-)-threo-3-Hydroxyaspartic acid	Creatinine	NADH/NAD+
2-Aminoadipic acid	Cystathionine	NADP_NADPH
2-Aminobutyric acid	Cysteine	N-alpha-Acetyllysine
2-Aminophenylacetic acid	d-11 hexanoic acid	Nervonic acid (C24_1n-

		9c)
2-Hydroxybutyric acid	d15-Octanoic acid	Nicotinamide
2-Hydroxycinnamic acid	d2-tyrosine	Nicotinic acid
2-Hydroxyglutaramic acid MCF1	d3- Alanine	Nonacosane
2-Hydroxyglutaramic acid MCF2	d4-Alanine	Nonadecanoic acid (C19_0)
2-Hydroxyisobutyric acid	d4-Citric Acid	Norleucine
2-Isopropylmalic acid	d-5-phenylalanine	Norvaline
2-ketoglutamarate	d5-Tryptophan	O-Acetylserine
2-Methyloctadecanoic acid	DBP	Octanoic acid (C8_0)
2-Oxoadipic acid	D-chiro-inositol Acetylated	Oleic acid (C18_1n-9c)
2-Oxobutyric acid	Decanoic acid (C10_0)	Ornithine
2-Oxoglutaric acid	Dehydroabietic acid	Oxalic acid
2-Oxovaleric acid	Dehydroascorbic acid	Oxaloacetic acid
2-Phosphoenolpyruvic acid	delta-Hydroxylysine	Palmitelaidic acid (C16_1n-9c)
2-Phosphoglyceric acid	DHA (C22_6n-3,6,9,12,15,18c)	Palmitic acid (C16_0)
3-(-2-Thienyl)-D-alanine	Dipicolinic acid	Palmitoleic acid (C16_1n-7c)
3,3-Dimethylglutaric acid	DL-3-Aminoisobutyric acid	para-Toluic acid
3,5-Diiodo-L-tyrosine	DL-gamma-methyl-ketoglutaramate isomer 1	Pentadecane
3-Acetoxy-3-Hydroxy-2-methylpropionic acid (NIST match: 88.6%)	DL-gamma-methyl-ketoglutaramate isomer 2	Pentadecanoic acid (C15_0)
3-Hydroxybenzoic acid	Dodecane	Phenethyl acetate
3-Hydroxydecanoic acid	Dodecanoic acid (C12_0)	Phenylalanine
3-Hydroxyoctanoic acid	DPA (C22_5n-3,6,9,12,15c)	Pimelic acid
3-Hydroxypropionic acid	D-stearic acid	Proline
3-Methyl-2-oxopentanoic acid	EDTA	Putrescine

3-Methyl-2-oxovaleric acid	EPA (C20_5n-3,6,9,12,15c)	Pyroglutamic acid
3-Oxoadipic acid	Erucic acid (C22_1n-9c)	Pyruvic acid
4 Hydroxybenzene sulphonic acid	Ferulic acid	S-Adenosylhomocysteine
4-Aminobenzoic acid	Fumaric acid	S-Adenosylmethionine
4-Aminobutyric acid (GABA)	gamma-Linolenic acid (C18_3n-6,9,12c)	Salicylic acid
4-Hydroxycinnamic acid	Glutamic acid	scyllo-inositol Acetylated
4-Hydroxyphenylacetic acid	Glutamine	Sebacic acid
4-Hydroxyphenylethanol	Glutaric acid	Serine
4-Methyl-2-oxopentanoic acid	Glutathione	Sinapic acid
5-Hydroxy-L-lysine	Glyceric acid	Stearic acid (C18_0)
5-Hydroxymethyl-2-furaldehyde	Glycerol	Suberic acid
5-Methoxytryptophan	Glycine	Succinic acid
5-Methyltryptophan	Glyoxylic acid	Syringic acid
9-Heptadecenoic acid (C17_1n-8t)	Gondoic acid (C20_1n-9c)	Tartaric acid
Adipic acid	Guanine	tert-Leucine
Adrenic acid (C22_4n-6,9,12,15c)	Heneicosanoic acid (C21_0)	Thiamine
Alanine	Heptadecane	Threonine
alpha-Linolenic acid (C18_3n-3,6,9c)	Hexanoic acid (C6_0)	trans-4-Hydroxyproline
Anthranilic acid	Hippuric acid	trans-Cinnamic acid
Arachidic acid (C20_0)	Histidine	trans-Vaccenic acid
Arachidonic acid (C20_4n-6,9,12,15c)	Homocysteine	Tricosane
Asparagine	Hydroxybenzoic acid	Tricosanoic acid (C23_0)
Aspartic acid	Indole-3-butyric acid	Tridecane
Azelaic acid	Isocitric acid	Tridecanoic acid (C13_0)

Behenic acid (C22_0)	Isocitric acid secondary peak	Tryptophan
Benzoic acid	Isoleucine	Tyrosine
Benzothiazole	Itaconic acid	Undecanoic acid (C11_0)
beta-Alanine	Lactic acid	Valine
beta-Citryl-L-glutamic acid	Leucine	Vanillic acid
beta-Methylamino-alanine (BMAA)		Levulinic acid
BHT (Antioxidant)		Lignoceric acid (C24_0)

Appendix 6. AMDIS settings used for the identification of metabolites from plasma and hair

		In-house library	NIST library	
		Plasma	**Hair**	**Plasma & hair**
Identification				
	Minimum match factor	60	70	34
	Type of analysis	Use Retention Time	Use Retention Time	Simple
	RT window	0.3 min	0.2 min	-
Instrument	Low m/z	38	38	38
	High m/z	550	550	550
	Scan direction	High to Low	High to Low	High to Low
	Data file format	Agilent files	Agilent files	Agilent files
	Instrument type	Quadrupole	Quadrupole	Quadrupole
Deconvolution	Component width	14	12	14
	Adjacent peak subtraction	One	Two	Two
	Resolution	Medium	High	Medium
	Sensitivity	Medium	Medium	Low
	Shape requirements	Low	Low	Low

Appendix 7. Metabolites identified in GUSTO plasma

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Leucine	In-house	100	61-90-5	13.752	Amino acid
Proline	In-house	100	147-85-3	14.729	Amino acid
Ethylenediaminetetraacetic acid	In-house	100	60-00-4	23.140	Carboxylic acid
Lactic acid	In-house	99	10326-41-7	8.949	Hydroxy acid
2-Hydroxybutyric acid	In-house	99	5094-24-6	10.143	Hydroxybutyric acid
Isoleucine	In-house	99	73-32-5	13.802	Amino acid
Glutamic acid	In-house	99	56-86-0	17.771	Amino acid
Palmitelaidic acid	In-house	99	10030-73-6	20.556	Fatty acid
Docosahexaenoic acid	In-house	99	6217-54-5	29.244	Fatty acid
Myristic acid	In-house	98	544-63-8	17.425	Fatty acid
Cysteine	In-house	98	52-90-4	19.581	Amino acid
Methionine	In-house	97	63-68-3	17.843	Amino acid
Stearic acid	In-house	97	57-11-4	23.922	Fatty acid
4-Methyl-2-oxopentanoic acid	In-house	96	4502-00-5	7.656	Keto acid
Citric acid	In-house	96	77-92-9	16.060	Carboxylic acid
Tryptophan	In-house	96	73-22-3	32.668	Amino acid
Alanine	In-house	95	56-41-7	10.796	Amino acid
Phenylalanine	In-house	95	63-91-2	19.464	Amino Acid
Margaric acid	In-house	95	506-12-7	22.831	Fatty acid
Palmitic acid	In-house	95	57-10-3	23.330	Fatty acid
Adrenic acid	In-house	95	28874-58-0	29.136	Fatty acid
3-Methyl-2-oxopentanoic acid	In-house	94	3715-31-9	7.515	Keto acid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Lauric acid	In-house	94	143-07-7	14.959	Fatty acid
Docosapentaenoic acid	In-house	94	24880-45-3	29.361	Fatty acid
Threonine	In-house	93	72-19-5	15.350	Amino acid
Creatinine	In-house	93	60-27-5	16.430	Alpha amino acid
Caffeine	In-house	93	58-08-2	23.244	Carboxylic acid
11,14-Eicosadienoic acid	In-house	93	2091-39-6	26.358	Fatty acid
Valine	In-house	92	72-18-4	12.498	Amino acid
cis-Aconitic acid	In-house	92	585-84-2	15.215	Carboxylic acid
Ornithine	In-house	92	70-26-8	24.421	Amino acid
Lysine	In-house	92	56-87-1	25.845	Amino acid
Tyrosine	In-house	92	60-18-4	28.292	Amino acid
Succinic acid	In-house	91	110-15-6	8.774	Dicarboxylic acid
gamma-Linolenic acid	In-house	91	506-26-3	23.962	Fatty acid
9-Heptadecenoic acid	In-house	90	10136-52-4	22.658	Fatty acid
4-Hydroxybenzenesulfonic acid	In-house	88	98-67-9	18.287	Organic acid
Glycine	In-house	87	56-40-6	11.145	Amino acid
Glutathione	In-house	87	70-18-8	18.452	Carboxylic acid
Pyruvic acid	In-house	86	127-17-3	7.104	Keto-acid
Dodecane	In-house	86	112-40-3	9.352	Alkane
Glyoxylic acid	In-house	85	298-12-4	10.922	Carboxylic acid
Decanoic acid	In-house	85	110-42-9	12.209	Fatty acid
Pentadecanoic acid	In-house	84	1002-84-2	18.868	Fatty acid
Palmitoyl ethanolamide	In-house	84	544-31-0	26.667	Fatty amide

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Histidine	In-house	83	71-00-1	26.628	Amino acid
2-Oxovaleric acid	In-house	81	1821-02-9	6.108	Keto-acid
Asparagine	In-house	81	70-47-3	16.261	Amino acid
Glutamine	In-house	81	56-85-9	24.518	Amino acid
Caproic acid	In-house	79	142-62-1	6.051	Fatty acid
Malic acid peak 2	In-house	79	6915-15-7	14.840	Dicarboxylic acid
Arachidic acid	In-house	78	506-30-9	26.518	Fatty acid
Malonic acid	In-house	76	141-82-2	7.162	Dicarboxylic acid
Serine	In-house	76	302-84-1	17.077	Amino acid
4-Hydroxyphenylacetic acid	In-house	76	156-38-7	18.593	Phenol
Itaconic acid	In-house	74	97-65-4	9.893	Organic acid
Glutaric acid	In-house	74	110-94-1	10.458	Dicarboxylic acid
Pyroglutamic acid	In-house	74	98-79-3	16.087	Carboxylic acid
Oxalic acid	In-house	73	144-62-7	5.814	Dicarboxylic acid
Tridecane	In-house	72	629-50-5	12.296	Alkane
Citric acid secondary peak	In-house	72	77-92-9	15.696	Carboxylic acid
Malic acid peak 1	In-house	70	97-67-6	11.066	Dicarboxylic acid
Myristoleic acid	In-house	70	544-64-9	17.398	Fatty acid
Adipic acid	In-house	69	124-04-9	12.102	Fatty acid
7,11-Hexadecadienal	In-house	69	849947-71-3	26.116	Carboxylic acid
para-Toluic acid	In-house	68	99-94-5	11.268	Benzoic acid
2-Hydroxyisobutyric acid	In-house	67	594-61-6	8.865	Hydroxy acid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Nonadecanoic acid	In-house	64	646-30-0	25.051	Fatty acid
Decane	In-house	Unknown	124-18-5	7.770	Alkane
Fumaric acid	In-house	Unknown	110-17-8	8.345	Dicarboxylic acid
2-Oxoglutaric acid	In-house	Unknown	328-50-7	13.675	Keto acid
Linoleic acid	In-house	Unknown	60-33-3	23.790	Fatty acid
Conjugated linoleic acid	In-house	Unknown	121250-47-3	24.935	Fatty acid
Arachidonic acid	In-house	Unknown	506-32-1	26.030	Fatty acid
bishomo-gamma-Linolenic acid	In-house	Unknown	1783-84-2	26.193	Fatty acid
Eicosapentaenoic acid	In-house	Unknown	10417-94-4	26.280	Fatty acid
11,14,17-eicosatrienoic acid	In-house	Unknown	2091-27-2	26.360	Fatty acid
Cholesterol	In-house	Unknown	57-88-5	43.000	Sterol
Unknown 045(100) 59(89.0) 74(67.8)	In-house	93	-	8.565	-
Unknown 128(100) 42(10.5) 129(6.8)	In-house	88	-	20.182	-
Caprylic acid	In-house	83	111-11-5	9.220	Fatty acid
Unknown 088(100) 115(48.9) 59(30.2)	In-house	82	-	12.795	-
Unknown 045(100) 59(93.0) 160(76.1)	In-house	79	-	16.009	-
trans-Vaccenic acid OR cis-Vaccenic acid	In-house	79	693-72-1	23.934	Fatty acid
Unknown 114(100) 147(31.9) 115(27.2)	In-house	72	-	22.554	-
Unknown 112(100) 43(60.3) 115(49.2)	In-house	70	-	10.211	-
Unknown 128(100)	In-house	69	-	20.182	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
139(21.1) 42(19.1)					
Unknown 088(100) 44(30.3) 59(19.6)	In-house	69	-	23.078	-
Unknown 071(100) 57(94.4) 85(72.9)	In-house	Unknown	-	13.585	-
n-Butylbenzene	NIST	97	104-51-8	7.885	Benzene
n-Pentylbenzene	NIST	96	538-68-1	9.494	Benzene
Cholesta-3,5-diene	NIST	96	747-90-0	36.433	Sterol
Cholesta-4,6-dien-3-one	NIST	95	566-93-8	46.392	Sterol
Cholesterol Chloride	NIST	92	910-31-6	35.413	Sterol
4,6-Cholestadien-3β-ol	NIST	91	14214-69-8	36.223	Sterol
Cholestenone	NIST	91	601-57-0	47.047	Sterol
1-methylindan	NIST	90	767-58-8	9.294	Indane
l-Leucine, N-methoxycarbonyl-, ethyl ester	NIST	90	88406-43-3	18.562	Amino acid
3-Hydroxy-2-methylpropanoic acid	NIST	89	80657-57-4	24.308	Hydroxy acid
Succinic acid, 2,2,3,3-tetrafluoropropyl 3-methylbut-2-yl ester	NIST	88		22.214	Dicarboxylic acid
3-Carboxy-4-methyl-5-propyl-2-furanpropionic acid, dimethyl ester	NIST	87	72719-10-9	19.649	Fatty acyl
Hexadecanoic acid, 15-methyl-, methyl ester (Palmitic acid peak 1)	NIST	87	6929-04-0	22.300	Fatty acid
l-Valine, N-methoxycarbonyl-, pentyl ester	NIST	86	1046143-00-3	17.575	Amino acid
Cholesterol, chlorodifluoroacetate	NIST	86	25952-12-9	35.413	Sterol

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
gamma-Tocopherol	NIST	84	2733-27-9	45.154	Vitamin
D-Alanine, N-ethoxycarbonyl-, heptyl ester	NIST	83		11.917	Amino acid
Glycine, N,N-bis(2-methoxy-2-oxoethyl)-, methyl ester	NIST	83	22241-07-2	16.819	Amino acid
Glycine, N-ethyl-N-(2-methoxyethoxycarbonyl)-, 2-methoxyethyl ester	NIST	82		14.932	Amino acid
Hexadecanoic acid, 14-methyl-, methyl ester (Palmitic acid peak 2)	NIST	82	2490-49-5	22.466	Fatty acid
5-Aminoimidazole-4-carboxylic acid, methyl ester	NIST	81	4919-00-0	18.469	Azole
1-Aminocyclopentanecarboxylic acid, N-methoxycarbonyl-, octyl ester	NIST	80	626-04-0	15.053	Carboxylic acid
tert-Butyldimethylsilylamine	NIST	78	41879-37-2	6.618	Amine
l-Alanine, N-methoxycarbonyl-, ethyl ester	NIST	78	88406-40-0	16.338	Amino acid
Tetradecanoic acid, 12-methyl-, methyl ester, (S)- (Myristic acid)	NIST	78	62691-05-8	18.432	Fatty acid
Indolepropionic acid, methyl ester	NIST	78	5548-09-4	23.412	Indole
3-nitropyrazole	NIST	76	26621-44-3	8.924	Azole
3-Fluoroanisidine	NIST	75	366-99-4	10.607	Aniline
1-Propene-1,2,3-tricarboxylic acid,	NIST	75	20820-77-3	18.938	Carboxylic acid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
trimethyl ester					
trans-2-Tetradecenal	NIST	75	51534-36-2	18.432	Fatty aldehyde
2-heptenylbenzene (Z)-	NIST	74	54725-18-7	13.828	Benzene
3,6,9,12-Tetraoxatridecanoic acid, ethyl ester	NIST	73	66664-66-2	14.661	Fatty acid
3-tert-Butyl-2-oxazolidinone	NIST	73	40482-46-0	16.568	Oxazolidinone
4,7,10-Hexadecatrienoic acid, methyl ester	NIST	73	17364-31-7	29.015	Fatty acid
l-Norvaline, N-(2-methoxyethoxycarbonyl)-, propyl ester	NIST	72		15.110	Amino acid
11-Eicosenoic acid, methyl ester, (11Z)-	NIST	72	2390-09-2	26.289	Fatty acid
Butanamide, N-(4-hydroxyphenyl)-	NIST	72	101-91-7	27.197	Fatty amide
Cholesta-3,5,24-triene	NIST	71	1036314-86-9	38.592	Sterol
1H,5H-benzo[ij]quinolizine, 9-[2-(9-anthracenyl)ethenyl]-2,3,6,7-tetrahydro-	NIST	70		14.519	
Ethyl 2,5,8,11,14-pentaoxahexadecan-16-oate	NIST	68		9.854	Fatty acid
Propylamine, 3-chloro-N,N-bis(2-methoxyethyl)-	NIST	68	959243-79-9	15.896	Amine
2-p-Tolylisoindole-1,3-dione	NIST	67	2142-03-2	30.520	
Methyldiethanolamine	NIST	66	105-59-9	9.860	Amine
Acetic acid, diethoxy-, ethyl ester	NIST	66	6065-82-3	10.934	Carboxylic acid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Acetamide, 2-anilino-N,N-dimethyl-	NIST	66	14307-89-2	16.308	Carboximidic acid
cis-4-Hydroxyproline	NIST	66	618-27-9	19.258	Carboxylic acid
3-Furancarboxamide, 2,5-dimethyl-N-phenyl-	NIST	66	28562-70-1	24.141	Carboxylic acid
10-Nonadecenoic acid, methyl ester	NIST	66	56599-83-8	24.868	Fatty acid
9-Octadecenoic acid (Z)-, 2-hydroxyethyl ester	NIST	66	4500-01-0	29.441	Fatty acid
Phenethyl acetate	NIST	65	103-45-7	12.302	Benzene
4'-Hydroxyvalerophenone	NIST	65	33342-92-6	15.320	Hydroxy ketone
Glycine, N,N-dipropyl-, ethyl ester	NIST	64	2644-22-6	18.706	Amino acid
2-Methyloctadecanoic acid	NIST	64	7217-83-6	24.643	Fatty acid
9-Octadecenamide, (Z)-	NIST	64	301-02-0	29.963	Fatty acid
p-Dioxane-2,3-diol	NIST	63	4845-50-5	5.644	
2-Bromomethyl-benzoic acid	NIST	63	7115-89-1	9.105	Benzoic acid
N-(2-Methoxyethyl)carbamic acid methyl ester	NIST	63	16339-52-9	14.121	Carboxylic acid
Palmitoyl chloride	NIST	63	112-67-4	26.667	Fatty acid
Succinic acid, 2,2,3,3,4,4,5,5-octafluoropentyl oct-3-en-2-yl ester	NIST	63		26.845	Dicarboxylic acid
Heptasiloxane, hexadecamethyl-	NIST	62	541-01-5	23.447	
4-Hexadecen-6-yne, (E)-	NIST	62	74744-51-7	26.224	

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Glycine, 2-cyclohexyl-N-(but-3-en-1-yl)oxycarbonyl-, octyl ester	NIST	61		26.027	Amino acid
Glycine, ethyl ester	NIST	61	459-73-4	28.513	Amino acid
7,7-Dimethyl-3,6,8,11-tetraoxa-7-silatridecane	NIST	60	17980-57-3	15.660	
Guanidine	NIST	59	80-70-6	9.706	
Bicyclo[2.2.1]heptan-2-amine, N,1,7,7-tetramethyl-	NIST	59	22285-82-1	16.191	
4-Iminobarbituric acid	NIST	59		19.182	
Benzene, hexyl-	NIST	59	1077-16-3	19.441	
1-Adamantanecarboxylic acid, morpholide	NIST	59	22508-50-5	28.716	
Butanoic acid, 4-nitro-	NIST	58	16488-43-0	10.658	
Glutaric acid, decyl 4-(trifluoromethyl)benzyl ester	NIST	58		15.829	
5-Octadecenoic acid, methyl ester	NIST	58	56554-45-1	23.768	
Glycine, N-(methoxyoxoacetyl)-, methyl ester	NIST	57	89464-63-1	16.804	
Methylenedioxyamphetamine acetate	NIST	57	36209-71-9	23.140	
Cholesteryl benzoate	NIST	57	604-32-0	36.334	
2-tert-Butyl-5,5-dimethyl-1-pyrroline, 1-oxide	NIST	56	58134-16-0	10.271	
Heptadecanoic acid	NIST	56	506-12-7	20.910	
1H-Indole-3-propanoic acid	NIST	56	830-96-6	23.412	

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Ethyl 6-methyl-3-oxopentadecanoate	NIST	56	959097-04-2	23.495	
1-Oxaspiro[2.5]octane, 5,5-dimethyl-4-(3-methyl-1,3-butadienyl)-	NIST	56	925919-79-5	26.368	
1,2,3-Trimethyl-cyclopent-2-enecarboxaldehyde	NIST	55	1228151-06-1	10.392	
Quinazoline, 4-(1-methylethyl)-, 1-oxide	NIST	55	50915-32-7	15.377	
6-Methyl-7-thioxo-4,7-dihydro-triazolo(3,2-c)triazine	NIST	55	62170-09-6	16.174	
Benz[d]isoxazole-5-ol-4-one, 4,5,6,7-tetrahydro-3-methyl-	NIST	55	959020-57-6	24.735	
o-Tolunitrile, .alpha.-cyano-	NIST	54	3759-28-2	15.002	
(-)-Neoclovene-(I), dihydro-	NIST	54	1005253-95-1	23.824	
1-(2,3-Dihydro-indol-1-yl)-2-(9-methyl-9H-1,3,4,9-tetraaza-fluoren-2-ylsulfanyl)-ethanone	NIST	53	352328-94-0	30.806	
1,2-Benzenediol, o-(4-methoxybezoyl)-o'-(5-chlorovaleryl)-	NIST	52		41.641	
2H-naphth[2,1-e]-1,3-oxazine-2,4(3H)-dione	NIST	51		11.876	
Hydrazine, (3-methoxyphenyl)-	NIST	51	15384-39-1	24.076	
Pyromellitic dianhydride	NIST	50	89-32-7	22.755	
Bitoscanate	NIST	50	4044-65-9	29.888	
8-Methoxy-1,3,4,4a,5,6-hexahydro-2H-	NIST	50	197661-85-1	31.215	

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
benzo[5,6]cyclohepta[1,2-b]pyridin-2-one					
7-Benzofuranamine, 2-methyl-	NIST	49	26325-21-3	15.462	
4-Hydroxybenzyl alcohol, 2TBDMS derivative (peak 1)	NIST	49	623-05-2	18.287	
4-Hydroxybenzyl alcohol, 2TBDMS derivative (peak 2)	NIST	49	623-05-2	21.125	
N-acetyl-3,4-dimethoxyamphetamine	NIST	49	51920-75-3	27.255	
5-Cyano-4-methoxyamino-7-phenyl-hept-6-enoic acid, methyl ester	NIST	48	959303-03-8	16.729	
Lycopodane, 4,5-didehydro-15-methyl-, (15R)-	NIST	48	54551-08-5	28.495	
1,2,3,4-Tetrahydroquinoline, N-methyl-	NIST	47	491-34-9	19.474	
3,4-methylenedioxyphenylpropan-2-ol, formyl	NIST	47	73549-48-1	19.496	
(7H)Thiopyrano[3,4-c]isoxazole, 3,3a,4,5-tetrahydro-7-(1-methylethyl)- (peak 1)	NIST	47	128869-51-2	21.283	
5-Oxotetrahydrofuran-2,3-dicarboxylic acid, dimethyl ester	NIST	47	16496-38-1	24.250	
Ethanol, 2-(dibutylamino)-	NIST	47	102-81-8	28.748	
(7H)Thiopyrano[3,4-c]isoxazole, 3,3a,4,5-tetrahydro-7-(1-	NIST	47	128869-51-2	21.283	

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
methylethyl)- (peak 2)					
1,2-Bis(3-aminopropylamino)ethane, N1,N2,N3,N4-tetrakis(trimethylsilyl)-	NIST	46		22.487	
3,3,3-Trifluoro-2-(4-hydroxy-2-methoxy-phenylsulfanyl)-2-trifluoromethyl-propionic acid methyl ester	NIST	45	137380-37-1	41.404	
Imidazo[4,5-e][1,4]diazepin-8(1H)-one, 4,5,6,7-tetrahydro-4,7-dimethyl-5-thioxo-	NIST	44	959006-96-3	26.636	
2-Ethyl-2-phenylaziridine	NIST	43	768-82-1	15.972	
Metacetamol	NIST	43	621-42-1	17.629	
1,1,4,4-Tetranitrocyclohexane	NIST	43	146028-82-2	23.370	
l-Alanyl-l-proline, N-methoxycarbonyl-, methyl ester	NIST	42	1419388-75-2	25.200	
2-Oxo-2,3-dihydro-1H-imidazole-4-carbonitrile	NIST	41	159263-03-3	5.944	
Androst-4-en-11-ol-3,17-dione, 9-thiocyanato-	NIST	41	1166-72-9	23.626	
1,2,3,4-Tetrahydrobenzimidazo[2,1-b]quinazolin-12-ol	NIST	41	24415-63-2	25.430	
3-Oxabicyclo[3.3.0]octan-2-one, 7,7-bis(methoxy)-	NIST	40	959247-51-9	15.181	
9-Octadecen-12-ynoic acid, methyl ester	NIST	40	56847-05-3	28.324	
Ethyl 8-amino-6-methoxycinchoninate	NIST	40	858466-65-6	32.797	
1H-1,3-Benzimidazol-4-	NIST	38		22.763	

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
amine, 5-methyl-					
1-Naphthaleneacetic acid, tridec-2-yn-1-yl ester	NIST	37		37.724	
4-Pyrimidinamine, 6-methyl-	NIST	36	3435-28-7	12.461	
Glutaric acid, 3-oxobut-2-yl propyl ester	NIST	36		18.913	
Cyclobutene, 2-propenylidene-	NIST	33	52097-85-5	10.899	
Pyrazole, 5-methyl-3-(5-nitro-2-furyl)-	NIST	31	16239-90-0	18.840	
6-Methyl-3,5-dithioxo-2,3,4,5-tetrahydro-1,2,4-dro-triazine	NIST	30	452-88-0	24.546	
2-Pentafluorophenylbenzothiazole	NIST	26	69200-85-7	20.274	
1,1-Dimethyl-2-[2-(methoxycarbonyl)prop-2-yl]hydrazine	NIST	16	97812-30-1	36.899	
beta-Alanine	NIST	Unknown	107-95-9	12.860	
1-Oxaspiro[2.5]octane, 5,5-dimethyl-4-(3-methyl-1,3-butadienyl)-	NIST	Unknown		26.368	

In-house identifications are labelled as the parent compound, NIST identifications are labelled as the derivative.

Unknown match factors are due to manual integration of compounds.

Compound class only listed for metabolites detected with a match factor > 60 %.

Appendix 8. Hair collection SOP

GUSTO Protocol for Hair Collection

(Version 03 August 2009)

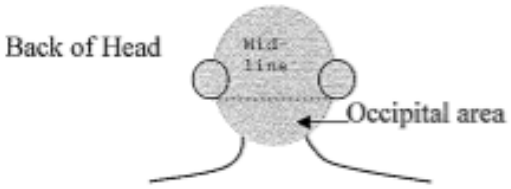
Purpose: For toxicology screening (exposure to lead, metals etc) and determine levels of drugs (e.g. steroids) in hair.

Specimen Requirements:

- Collect hair sample from pregnant mothers at 26 week clinic visit
- Collect approximately 5 strands of hair

Preparation

- Obtain preprinted labels for the hair sample.
- Put on new gloves.
- Use a comb to partition the hair between ears on the back of the head below the midline. This is the occipital area at the rear base of the head.



Procedure Summary

- Cut the hair approximately 0.5cm away from the scalp with a blunt scissor. Around the occipital area.
  Note:
  - *Collection of fallen hairs after the participants comb their hair is also acceptable.*
  - *Please note that it does not need to retain the orientation of the hair strands.*
- Twist the hair together
- Place the hair sample into a zip closable bag
  Note:
  - *Immediately close and seal the zip closable bag*
  - *Do not place any hair that has been fallen to the floor into the bag*
  - *Please make sure hairs are not protruding through the opening*
- Label the zip closable bag with the preprinted label. Store the hair at room temperature
- Transport specimen to the laboratory
- Store the hair sample at -20°C at the laboratory.

Appendix 9. Metabolites identified in GUSTO hair samples

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Unknown 201(100) 117(89.6) 119(86.2)	In-house	99	-	8.164	-
Glycine	In-house	99	56-40-6	11.450	Amino acid
Leucine	In-house	99	61-90-5	14.013	Amino acid
Proline	In-house	99	147-85-3	15.006	Amino acid
Phenylalanine	In-house	99	63-91-2	19.935	Amino acid
Unknown 127(100) 159(49.8) 59(48.8)	In-house	98	-	10.514	-
Threonine	In-house	98	72-19-5	15.646	Amino acid
Glutamic acid	In-house	98	56-86-0	18.122	Amino acid
10,13-dimethyltetradecanoic acid	In-house	98	267650-23-7	21.276	Fatty acid
Oxalic acid	In-house	97	144-62-7	5.972	Dicarboxylic acid
Alanine	In-house	97	56-41-7	11.054	Amino acid
Isoleucine	In-house	97	73-32-5	14.063	Amino acid
Serine	In-house	97	302-84-1	17.382	Amino acid
Ornithine	In-house	97	70-26-8	24.797	Amino acid
Pyruvic acid	In-house	96	127-17-3	7.302	Keto acid
Benzoic acid	In-house	96	26818-10-0	9.603	Benzoic acid
Unknown 144(100) 88(51.5) 115(46.4)	In-house	96	-	13.953	-
Aspartic acid	In-house	96	56-84-8	16.319	Amino acid
Tyrosine	In-house	96	60-18-4	28.680	Amino acid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Unknown 277(100) 278(36.9) 199(14.0)	In-house	96	-	35.234	-
2-Oxobutyric acid	In-house	95	600-18-0	5.802	Keto acid
Malonic acid	In-house	95	141-82-2	7.387	Dicarboxylic acid
4-Methyl-2-oxopentanoic acid	In-house	95	4502-00-5	7.876	Keto acid
Unknown 115(100) 118(95.4) 86(69.3)	In-house	95	-	15.750	-
Dipicolinic acid	In-house	95	499-83-2	18.305	Pyridine-carboxylic acids
Unknown 086(100) 59(56.6) 128(39.3)	In-house	94	-	13.195	-
Unknown 128(100) 42(0.90) 129(0.88)	In-house	94	-	24.634	-
Unknown 074(100) 46(21.3) 59(0.97)	In-house	94	-	6.820	-
Citraconic acid	In-house	94	498-23-7	10.213	Methyl-branched fatty acid
cis-Aconitic acid	In-house	94	585-84-2	15.486	Tricarboxylic acid
Pyroglutamic acid	In-house	94	98-79-3	16.423	Carboxylic acid
Cysteine	In-house	94	52-90-4	20.085	Amino acid
Unknown 128(100) 139(21.1) 42(19.1)	In-house	93	-	19.475	-
Succinic acid	In-house	93	110-15-6	9.023	Dicarboxylic acid
Glyoxylic acid	In-house	93	298-12-4	11.173	Carboxylic acid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Valine	In-house	93	72-18-4	12.755	Amino acid
Stearic acid (C18_0)	In-house	93	57-11-4	24.149	Fatty acid
Unknown 135(100) 180(28.0) 77(19.9)	In-house	92	-	15.830	-
Unknown 114(100) 59(61.6) 146(33.7)	In-house	92	-	17.06	-
2-Oxovaleric acid	In-house	92	1821-02-9	6.262	Keto acid
Unknown 070(100) 42(28.5) 69(8.0)	In-house	92	-	6.865	-
Lactic acid	In-house	92	10326-41-7	9.185	Hydroxy acid
2,4-Di-tert-butylphenol	In-house	92	96-76-4	16.368	Benzenoid
Unknown 112(100) 43(60.3) 115(49.2)	In-house	91	-	10.449	-
Unknown 056(100) 59(24.8) 115(10.8)	In-house	91	-	6.916	-
Adipic acid	In-house	91	124-04-9	12.349	Fatty acid
Malic acid peak 2	In-house	91	6915-15-7	15.118	Dicarboxylic acid
Unknown 091(100) 150(35.5) 65(11.2)	In-house	90	-	11.156	-
Unknown 045(100) 59(89.0) 74(67.8)	In-house	90	-	8.910	-
Unknown 174(100) 70(96.8) 114(77.2)	In-house	90	-	30.772	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
3-Methyl-2-oxopentanoic acid	In-house	89	3715-31-9	7.733	Keto acid
para-Toluic acid	In-house	89	99-94-5	11.515	Benzenoid
Creatinine	In-house	89	60-27-5	16.648	Alpha amino acid
Glutathione	In-house	89	70-18-8	18.870	Carboxylic acid
Levulinic acid	In-house	88	123-76-2	8.660	Keto acid
Unknown 126(100) 127(46.7) 59(33.2)	In-house	88	-	9.932	-
2-Aminobutyric acid	In-house	88	2623-91-8	12.172	Alpha amino acid
2-Oxoglutaric acid	In-house	88	328-50-7	13.061	Keto acid
Myristic acid (C14_0)	In-house	88	544-63-8	17.684	Fatty acid
Citramalic acid	In-house	87	2306-22-1	10.670	Hydroxy fatty acid
4-Hydroxyphenylacetic acid	In-house	87	156-38-7	18.982	Benzenoid
Dibutyl phthalate	In-house	87	84-74-2	15.948	Benzenoid
Unknown 130(100) 70(47.4) 98(19.7)	In-house	86	-	24.691	-
Unknown 082(100) 142(56.8) 110(42.6)	In-house	86	-	14.933	-
Decanoic acid	In-house	86	110-42-9	12.468	Fatty acid
Citric acid	In-house	86	77-92-9	16.345	Amino acid
Palmitelaidic acid or Palmitoleic acid	In-house	86	10030-73-6	20.922	Fatty acid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Fumaric acid	In-house	85	110-17-8	9.157	Dicarboxylic acid
Dodecanoic acid	In-house	85	143-07-7	15.186	Fatty acid
Lysine	In-house	84	56-87-1	26.278	Amino acid
Dehydroabietic acid	In-house	84	1235-74-1	28.381	Diterpenoids
Glutaric acid	In-house	83	110-94-1	9.417	Dicarboxylic acid
Methionine	In-house	83	63-68-3	18.193	Amino acid
Margaric acid	In-house	81	506-12-7	23.059	Fatty acid
Nicotinic acid	In-house	80	59-67-6	10.331	Pyridine-carboxylic acid
Pentadecanoic acid	In-house	80	1002-84-2	19.213	Fatty acid
10-Heptadecenoic acid or 9-Heptadecenoic acid	In-house	80	62528-80-7	22.866	Fatty acid
Tryptophan	In-house	80	73-22-3	33.355	Amino acid
Unknown 121(100) 192(52.8) 165(50.7)	In-house	79	-	17.039	-
Hydroxybenzoic acid	In-house	79	69-72-7	17.172	Benzenoid
Unknown 070(100) 239(38.0) 128(13.0) (split peak 1)	In-house	77	-	28.965	-
Unknown 102(100) 103(26.6) 44(19.0) (split peak 1)	In-house	73	-	20.350	-
Methyl 4-tert-	NIST	96	26537-19-9	14.740	Benzenoid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
butylbenzoate					
d-Prolyl-d-proline, N-methoxycarbonyl-, methyl ester	NIST	95	-	28.729	Peptide
2,4-dinitrophenyl-L-isoleucine methyl ester	NIST	94	85993-45-9	29.603	Amino acid derivative
Decane, 1-chloro-	NIST	93	1002-69-3	14.263	Fatty acid
Methyl pyrrole-2-carboxylate	NIST	93	1193-62-0	11.219	Substituted pyrrole
Diisooctyl phthalate	NIST	92	131-20-4	31.844	Benzenoid
Glycyl-l-proline, N-methoxycarbonyl-, methyl ester	NIST	91	1571138-85-6	26.116	Peptide
l-Leucine, N-methoxycarbonyl-, pentyl ester	NIST	91	1046149-34-1	25.822	Amino acid
l-Prolylglycine, N-methoxycarbonyl-, methyl ester (peak 1)	NIST	91	-	23.971	Peptide
l-Prolylglycine, N-methoxycarbonyl-, methyl ester (peak 2)	NIST	91	-	25.510	Peptide
2,5-Piperazinedione	NIST	90	106-57-0	12.113	Piperazine
Dimethyl phthalate	NIST	90	131-11-3	15.948	Benzenoid
l-Alanyl-l-proline, N-methoxycarbonyl-, methyl ester	NIST	90	1419388-75-2	24.224	Peptide

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
l-Leucine, N-methyl-N-(2-methoxyethoxycarbonyl)-, hexyl ester	NIST	90	-	14.241	Amino acid
11-Octadecenoic acid, methyl ester	NIST	88	693-72-1	23.958	Fatty acid
1-Chloroundecane	NIST	88	2473-03-2	16.85	Alkane
5-Isopropyl-2,4-imidazolidinedione	NIST	88	16935-34-5	18.637	Ureide
Dodecane, 1-chloro-	NIST	88	112-52-7	16.866	Alkane
Glycyl-l-leucine, N-methoxycarbonyl-, methyl ester	NIST	88	204864-61-9	24.593	Peptide
Alanine, N-methyl-N-(2-methoxyethoxycarbonyl)-, nonyl ester	NIST	87	-	14.650	Amino acid
l-Leucylglycine, N-methoxycarbonyl-, methyl ester	NIST	87	-	23.995	Peptide
l-Leucyl-l-alanine, N-methoxycarbonyl-, methyl ester	NIST	87	1235348-16-9	23.361	Peptide
Benzonitrile, 2-methyl-	NIST	85	529-19-1	11.599	Benzenoid
d-Proline, N-methoxycarbonyl-, pentyl ester	NIST	84	1046139-78-9	23.562	Amino acid
Isobutyl methyl phthalate	NIST	83	73513-54-9	24.453	Benzenoid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Phthalic acid, 2-ethoxyethyl ethyl ester	NIST	82	13401-73-5	17.615	Benzenoid
1-Tetradecanol	NIST	81	112-72-1	17.534	Fatty alcohol
2,4-Imidazolidinedione, 5-(2-methylpropyl)-, (S)-	NIST	81	40856-75-5	20.603	Imidazolidine
D-Norleucine, N-ethoxycarbonyl-, ethyl ester	NIST	81	70288-79-8	15.323	Amino acid
l-Leucine, N-methyl-N-(2-methoxyethoxycarbonyl)-, undecyl ester (peak 1)	NIST	81	-	16.600	Amino acid
l-Leucine, N-methyl-N-(2-methoxyethoxycarbonyl)-, undecyl ester (peak 2)	NIST	81	-	17.798	Amino acid
l-Leucine, N-methyl-N-(2-methoxyethoxycarbonyl)-, undecyl ester (peak 3)	NIST	81	-	17.855	Amino acid
Morpholin-4-yl-acetic acid, hydrazide	NIST	81	770-17-2	17.984	Carboxylic acid
4-Aminobutyric acid (GABA)	NIST	81	56-12-2	14.860	Carboxylic acid
1,2-Benzenedicarboxylic acid	NIST	80	88-99-3	14.944	Benzenoid
Benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-	NIST	79	6386-38-5	23.034	Benzenoid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
hydroxy-, methyl ester					
l-Isoleucylglycine, N-methoxycarbonyl-, methyl ester	NIST	79	-	24.106	Peptide
Nonanoic acid, methyl ester	NIST	79	1731-84-6	10.993	Fatty acid
Carbamic acid, N-(1-methylcyclopropyl)-, methyl ester	NIST	78	104131-52-4	9.627	Carbamic acid
Dimethyl aminomalonic acid	NIST	78	16115-80-3	14.981	Carboxylic acid
1-Dodecanol	NIST	77	112-53-8	14.990	Fatty alcohol
d-Proline, N-methoxycarbonyl-, isobutyl ester	NIST	77	-	38.779	Amino acid
Dimethyl fumarate	NIST	77	23055-10-9	8.592	Fatty acid ester
Benzeneacetic acid, 3,4-dihydroxy-	NIST	76	102-32-9	17.501	Benzenoid
L-Phenylalanine, N-acetyl-, methyl ester	NIST	76	3618-96-0	31.244	Amino acid
dl-2-Aminocaprylic acid	NIST	75	644-90-6	13.565	Carboxylic acid
l-Norvaline, N-ethoxycarbonyl-, heptyl ester	NIST	75	-	19.967	Amino acid
l-Valine, N-methoxycarbonyl-, isohexyl ester	NIST	75	1046142-97-5	23.207	Amino acid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
N,N-Diethylhexylamine	NIST	75	44979-90-0	11.964	Aralkylamine
Carbonic acid, dodecyl methyl ester	NIST	74	6222-09-9	17.632	Carbonic acid
l-Leucyl-l-isoleucine, N-methoxycarbonyl-, methyl ester (peak 1)	NIST	74	700819-77-8	25.119	Peptide
l-Leucyl-l-isoleucine, N-methoxycarbonyl-, methyl ester (peak 2)	NIST	74	700819-77-8	25.119	Peptide
Pyrovalerone Hydrochloride	NIST	74	1147-62-2	16.503	-
Methyl acetoacetate	NIST	74	105-45-3	6.897	Keto acid
l-Allylglycine, N-ethoxycarbonyl-, pentyl ester	NIST	73	1000325-15-3	17.369	Peptide
l-Leucine, N-(2-methoxyethoxycarbonyl)-, octyl ester	NIST	73	-	15.629	Amino acid
l-Prolylglycine, N-methoxycarbonyl-, 2,2,2-trifluoroethyl ester	NIST	72	-	36.697	Peptide
Alanine, N-methyl-N-(2-chloroethoxycarbonyl)-, 2-chloroethyl ester	NIST	71	-	14.614	Amino acid
l-Valine, N-methoxycarbonyl-	NIST	71	1046143-00-3	22.637	Amino acid

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
, pentyl ester					
5-Thiazoleethanol, 4-methyl-	NIST	70	137-00-8	12.690	Thiazole
d-Proline, N-methoxycarbonyl-, hexyl ester	NIST	70	1046124-95-1	40.658	Amino acid
l-Alanyl-l-leucine, N-methoxycarbonyl-, methyl ester	NIST	70	-	23.502	Peptide
l-Proline, N-methoxycarbonyl-, butyl ester	NIST	70	1046125-41-0	30.566	Amino acid
2-Phenylindolizine	NIST	69	25379-20-8	22.778	Indolizine
Butanoic acid, .alpha.-[(2-methoxy-1,2-dioxoethyl)hydrazono]-	NIST	69	959294-51-0	20.749	Fatty acid
N-Ethyl-5-methyl-5-undecanamine	NIST	69	71275-05-3	17.083	
3-Pyridinecarboxamide, 1-oxide	NIST	68	1986-81-8	25.681	Nicotinamide
Naphthalene, 1-isothiocyanato-	NIST	67	551-06-4	22.969	Napthalene
1H-Tetrazaborole, 4,5-dihydro-1,4,5-trimethyl-	NIST	66	20546-18-3	23.443	
4-Ethyl-5-methylthiazole	NIST	66	52414-91-2	9.271	Thiazole
4-Vinyl-imidazole	NIST	66	3718-04-5	11.736	Imidazole
5,10-Diethoxy-2,3,7,8-tetrahydro-1H,6H-dipyrrolo[1,2-	NIST	66	850040-22-1	25.320	Pyrazine

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
a:1',2'-d]pyrazine					
Phenol, 3,5-dimethoxy- (split peak 1)	NIST	66	500-99-2	23.787	Phenol
Phenol, 3,5-dimethoxy- (split peak 2)	NIST	66	500-99-2	24.532	Phenol
Pyrrolo[1,2-a]pyrazine-3-propanamide, 2,3,6,7,8,8a-hexahydro-1,4-dioxo-	NIST	66	308355-86-4	34.662	Pyrazine
l-Norleucine, N-ethoxycarbonyl-, ethyl ester	NIST	65	70288-79-8	19.137	Amino acid
l-Valine, N-methoxycarbonyl-, hexyl ester (peak 1)	NIST	65	1046143-08-1	22.218	Amino acid
l-Valine, N-methoxycarbonyl-, pentyl ester (peak 2)	NIST	65	1046143-08-1	22.637	Amino acid
Ethanamine, 2-methoxy-N-(2-methoxyethyl)-N-methyl-	NIST	64	92260-33-8	22.673	Amine
Methyl isopropylcarbamate	NIST	64	5602-90-4	12.276	Benzenoid
Pyrrolo[1,2-a]pyrazine-1,4-dione, hexahydro-3-(2-methylpropyl)-	NIST	64	56548-64-3	24.543	Carboxylic acid
5-Aminoimidazole-4-carboxylic acid,	NIST	63	4919-00-0	8.442	Aminoimidazole

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
methyl ester					
9-Octadecenamide, (Z)-	NIST	63	301-02-0	40.400	Fatty acid
Benzenepropanoic acid, .alpha.-hydroxy-, methyl ester	NIST	63	13674-16-3	16.477	Fatty acid ester
Bicyclo[2.2.1]heptan-2-amine, N,1,7,7-tetramethyl-	NIST	63	22285-82-1	16.453	-
Carbamic acid, phenyl-, methyl ester	NIST	63	2603-10-3	20.657	Benzenoid
Fumaric acid, butyl pent-4-en-2-yl ester	NIST	63	-	21.675	Carboxylic acid
l-Proline, N-methoxycarbonyl-, tridecyl ester	NIST	63	-	41.127	Amino acid
4-Amino-2,6-dihydroxypyrimidine	NIST	63	873-83-6	6.124	Pyrimidine
3H-1,2,4-Triazole-3-thione, 5-amino-1,2-dihydro-	NIST	63	16691-43-3	6.309	Azole
1-Methylpyridin(2H)-2-one-5-carboxylic acid, methyl ester	NIST	62	6375-89-9	23.394	Carboxylic acid
Benzoic acid, 2-amino-5-chloro-, methyl ester	NIST	62	5202-89-1	17.105	Benzenoid
Diethyl 2,2'-(2,2'-oxybis(ethane-	NIST	62	-	10.105	

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
2,1-diyl)bis(oxy))diacetate					
Ethane, 1,1,1-trimethoxy-	NIST	62	1445-45-0	15.877	Glucoside
Ethanol, 2-(cyclohexylamino)-	NIST	62	2842-38-8	9.980	-
Octyl (E)-2-methylbut-2-enoate	NIST	62	29781-03-1	18.025	Fatty alcohol ester
5-Azacytosine (peak 1)	NIST	61	931-86-2	22.617	Pyrimidine
5-Azacytosine (peak 2)	NIST	61	931-86-2	29.667	Pyrimidine
Ethanedioic acid, dihydrazide	NIST	61	996-98-5	9.188	Dicarboxylic acid
l-Leucine, N-methoxycarbonyl-, pentyl ester	NIST	61	-	25.822	Amino acid
l-Norvaline, N-methoxycarbonyl-, hexyl ester	NIST	61	237076-57-2	24.313	Amino acid
Propanoic acid, 2-methyl-, decyl ester	NIST	61	5454-22-8	17.633	Carboxylic acid
1-Methyl -3,4-dihydro-2H-quinolin-7-amine	NIST	59	304690-94-6	23.413	-
2-Butenoic acid, 2-methyl-, 2-methylpropyl ester, (E)-	NIST	59	61692-84-0	18.050	-
Acetamide, N-acetyl-N-(1-methylethyl)-	NIST	59	1563-85-5	13.324	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Glycine, N-(2-methoxy-2-oxoethyl)-, methyl ester	NIST	59	6096-81-7	19.135	
Methane, diethoxy-	NIST	59	462-95-3	14.080	-
Silane, ethoxytriethyl-	NIST	59	597-67-1	8.552	-
3,4-Oxazolidinecarboxylic acid, 2-(1,1-dimethylethyl)-, dimethyl ester (2R-cis)-	NIST	58	116842-10-5	14.402	-
2-Amino-6,8-dihydroxypurine	NIST	57	5614-64-2	22.524	-
Carbonic acid, monoamide, N-(4-phenoxyphenyl)-, propyl ester	NIST	57	-	23.885	-
Ethyl orthoformate	NIST	57	-	11.156	-
1-Methoxyethanimine, N-acetyl-	NIST	56	99028-43-0	12.592	-
3H-Benzo[4,5]furo[3,2-d]pyrimidine-4-thione	NIST	56	62208-70-2	17.577	-
2,4,6,8-Tetrathiatricyclo[3.3.1.1(3,7)]decane, 9,10-dimethyl-, (.+/-.)-	NIST	55	17879-05-9	42.705	-
cis-1,3-Diacetamido-1,2,3,4-tetrahydronaphth	NIST	55	6580-54-7	25.181	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
alene					
l-Leucine, N-methoxycarbonyl-, isohexyl ester	NIST	55	1046135-10-7	14.848	-
Methanol	NIST	55	1825-61-2	16.634	-
2-Aminoimidazole-4-carboxylic acid, methyl ester	NIST	55	897920-30-8	8.442	-
[1,1'-Biphenyl]-2,2'-diol	NIST	54	1806-29-7	19.693	-
4-Amino-3-ethyl-2-thioxo-2,3-dihydro-thiazole-5-carboxylic acid methylamide	NIST	53	421586-27-8	22.456	-
4-Methoxyphenoxy phenylacetamide	NIST	53	6343-93-7	16.490	-
Benzene, [(1-methylpropyl)thio]-	NIST	53	14905-79-4	23.838	-
DL-Alanine, N-methyl-N-(vinyloxycarbonyl) -, hexyl ester	NIST	53	-	25.646	-
l-Phenylalanine, N-butyryl-, methyl ester	NIST	53	874290-39-8	29.831	-
3,4-Dimethoxybenzal dehyde oxime	NIST	52	93-04-9	15.324	-
Fumaric acid, ethyl 2-formylphenyl ester	NIST	52	171526-99-1	16.704	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
l-Isoleucylglycine, N-methoxycarbonyl-, 2,2,3,3,3-pentafluoropropyl ester	NIST	52	-	29.856	-
Butanamide, N-(4-hydroxyphenyl)-	NIST	51	101-91-7	27.715	-
d-Allylglycine, N-butoxycarbonyl-, ethyl ester	NIST	51	-	16.618	-
Methanamine, N-(5-phenyl-1,3,4-oxadiazol-2-yl)methyl-	NIST	51	880361-90-0	24.889	-
Methane, bromotrichloro-	NIST	51	75-62-7	9.563	-
p-Chloroaniline	NIST	51	106-47-8	14.456	-
Pyrimidine, 2-(4-butylphenyl)-5-ethyl-	NIST	51	116556-49-1	19.529	-
1H-Indole, 3-[(4,5-dihydro-3,5,5-trimethyl-1H-pyrazol-1-yl)methyl]-	NIST	50	767300-57-2	23.482	-
Glycine, N-(methoxyoxoacetyl)-, methyl ester	NIST	50	89464-63-1	17.098	-
Naphtho[2,1-b]furan, 1,2-dimethyl-	NIST	50	129812-23-3	25.401	-
[1,3,5]Triazine-2,4-diamine, 6-methoxy-N,N,N',N'-tetramethyl- (peak 1)	NIST	49	7710-30-7	15.056	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
[1,3,5]Triazine-2,4-diamine, 6-methoxy-N,N,N',N'-tetramethyl- (peak 2)	NIST	49	7710-30-7	18.940	-
2-Amino-5-methylamino-1,3,4-thiadiazole	NIST	48	33151-04-1	26.910	-
Benz[d]isoxazole-5-ol-4-one, 4,5,6,7-tetrahydro-3-methyl-	NIST	48	959020-57-6	25.093	-
1-(Naphth-1-ylamino)-2-phenyl-2-(1,2,4-thiazol-5-yl)ethene	NIST	47	153333-56-3	34.562	-
1,2,3,6-Tetrahydropyridine	NIST	47	694-05-3	9.500	-
2-Isopropyl-4-methoxy-oxazolidine-3-carboxylic acid, methyl ester (peak 1)	NIST	47	959020-91-8	21.838	-
2-Isopropyl-4-methoxy-oxazolidine-3-carboxylic acid, methyl ester (peak 2)	NIST	47	959020-91-8	29.063	-
Isoquinoline, 3-butyl-	NIST	47	7661-42-9	18.678	-
N(1)-(3-Methyl-1,2,4-oxadiazol-5-yl)-1-pyrrolidinecarbox	NIST	47	114212-83-8	23.540	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
amidine					
2-Chlorobenzalmalononitrile	NIST	46	2698-41-1	24.058	-
4-Chloro-2-methoxybenzyl alcohol, methyl ether	NIST	46	1822768-87-5	16.146	-
4-Methylanthranilic acid, N-methyl-N-nitroso-, methyl ester	NIST	46	959005-25-5	30.227	-
Cyclopentanetridecanoic acid, methyl ester	NIST	46	24828-61-3	22.744	-
Ethanone, 1,1'-(1,4-dihydro-2,4,6-trimethyl-3,5-pyridinediyl)bis-	NIST	46	1081-09-0	19.286	-
Glycine, N-methyl-n-propoxycarbonyl-, ethyl ester	NIST	46	1862712-00-2	24.366	-
2,4-Hexadienedioic acid, 3-methyl-4-propyl-, dimethyl ester, (E,Z)-	NIST	45	58367-44-5	22.945	-
3,4-Furandicarboxylic acid, diethyl ester	NIST	45	30614-77-8	29.472	-
Glycine, N-methyl-N-methoxycarbonyl-, heptyl ester	NIST	45	213267-07-3	22.933	-
1,4-naphthalenedione	NIST	44	7473-18-9	31.805	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
, 2-ethoxy-					
Benzene, 1-methoxy-4-(phenylmethyl)-	NIST	44	834-14-0	21.363	-
Decanoic acid, 2,4,6-trimethyl-, methyl ester	NIST	44	55955-72-1	21.809	-
Nanofin	NIST	44	504-03-0	7.485	-
2-(1-Hydroxy-1-methylethyl)pyrrolidine-1-carboxylic acid, methyl ester	NIST	43	959022-60-7	22.988	-
4-Iminobarbituric acid	NIST	43	14436-34-1	18.791	-
Benzene, (2-methyl-3-butenyl)-	NIST	43	1647-06-9	30.538	-
Propargylamine, N-trimethylsilyl-	NIST	43	1063798-95-7	12.920	-
1,3-Triazetidinedicarboxylic acid, 2-ethyl-4-methyl-, dimethyl ester	NIST	42	55131-06-1	22.404	-
2-Methoxy-5-nitro-biphenyl	NIST	42	15854-75-8	29.983	-
3-(4,7-Dimethoxy-2H-1,3-benzodioxol-5-yl)prop-2-enoic acid	NIST	42	74066-25-4	23.832	-
Butanamine, N,N-bis(2-hydroxypropyl)-	NIST	42	4402-34-0	26.494	-
Methyl 4-methoxy-4,8,12-trimethyltridecan	NIST	42	959088-55-2	13.800	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
oate					
Naphthalene, 1-methyl-	NIST	42	90-12-0	25.207	-
Naphthalene, 2-methoxy-	NIST	42	93-04-9	17.285	-
.delta.2-Tetrazaboroline, 5-ethyl-1,4-dimethyl-	NIST	41	20534-01-4	14.395	-
Benzaldehyde, 2,6-difluoro-3-methoxy-	NIST	41	149949-30-4	13.162	-
Benzonitrile, 4-[(methoxyimino) methyl]-	NIST	41	33499-34-2	27.372	-
Glutaric acid, cyclohexylmethyl 1,1,1-trifluoroprop-2-yl ester	NIST	41	-	24.020	-
Glycine, N-methyl-N-methoxycarbonyl-, heptyl ester	NIST	41	213267-07-3	28.225	-
3-Dibenzofuranamine	NIST	40	4106-66-5	23.897	-
3-Aminodihydro-2(3H)-furanone	NIST	39	1192-20-7	13.613	-
4-Methoxy-6-methylhexahydropyrimidin-2-thione	NIST	39	97481-97-5	17.157	-
Thiophen-2-methylamine, N-(2-fluorophenyl)-	NIST	39	886791-53-3	20.184	-

Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
Acetamide, 2-[2-(hexyloxy)ethoxy]-N,N-dipropyl-	NIST	38	77611-57-5	27.620	-
Benzoic acid, 3,4,5-trihydroxy-, 1-methyl-2-oxopropyl ester	NIST	38	26818-10-0	28.549	-
N,4-Diethyl-4-heptanamine	NIST	38	71275-07-5	35.092	-
3-Spirocyclohexane-1,2,3,4-tetrahydroisoquinolin-1-one, hydrazone	NIST	37	352447-65-5	23.215	-
L-Cysteine, ethyl ester	NIST	37	3411-58-3	15.238	-
2-Methyl-1,6-dihydro-4-methylamino-6-pyrimidinone	NIST	35	857414-73-4	15.100	-
Benzene, 1-ethoxy-4-fluoro-	NIST	35	459-26-7	23.054	-
Carbon Tetrachloride	NIST	35	56-23-5	13.583	-
Methyl 2-O-acetyl-3,4-di-O-methyl-6-deoxy-.alpha.-D-mannopyranoside	NIST	35	72922-26-0	23.633	-
5-Isothiazolemethanol	NIST	34	1710-66-3	15.778	-
alpha.-D-Galactopyranuronic acid, 1,2:3,4-bis-O-(1-methylethylidene)	NIST	32	25253-46-7	31.553	-

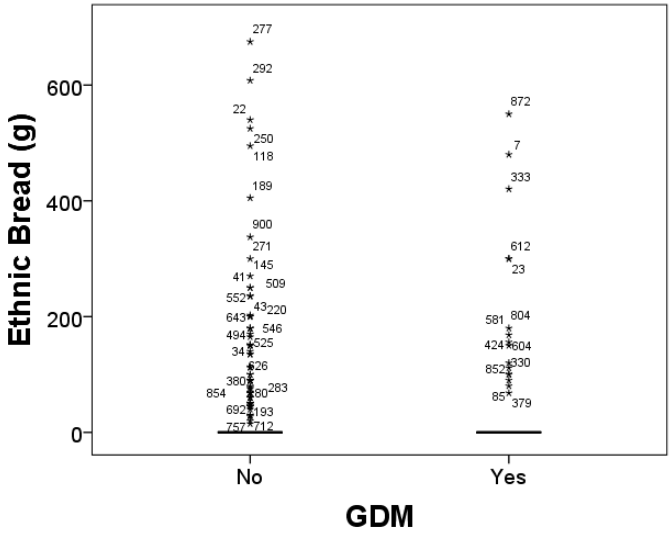
Metabolite	Library derived from	Match factor (%)	CAS number	Retention time (minutes)	Class
-					
1,4-Oxathiepan-2-one, 3,3-dimethyl-	NIST	32	35562-79-9	27.086	-
1-(4-Trimethylsilyloxycarbonylphenyl)-2-(pyrrolidin-1-yl)-butan-1-one	NIST	31	864444-69-9	29.327	-
4-Isopropylphenylcarbamic acid, methyl ester	NIST	30	133144-33-9	32.907	-
Dipyrrolo[1,2-a:-1',2'-d]pyrazine-2,5,10-trione, octahydro-7-hydroxy-	NIST	28	-	25.869	-
Nonylamine, 1-heptyl-	NIST	28	24539-83-1	29.355	-

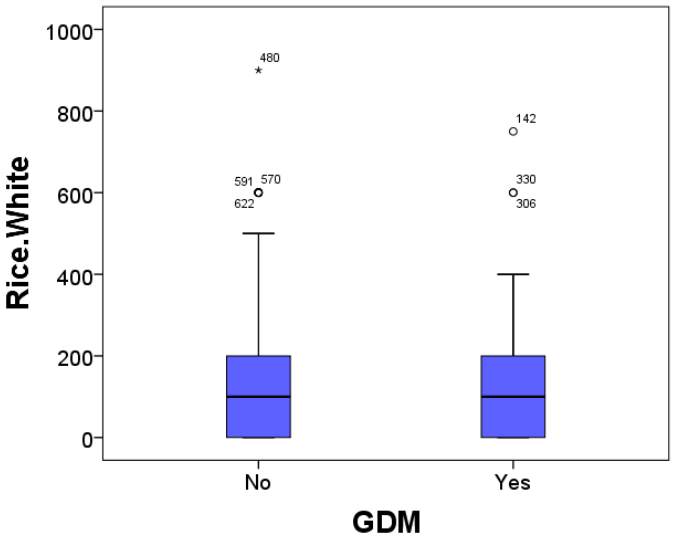
In-house identifications are labelled as the parent compound, NIST identifications are labelled as the derivative.

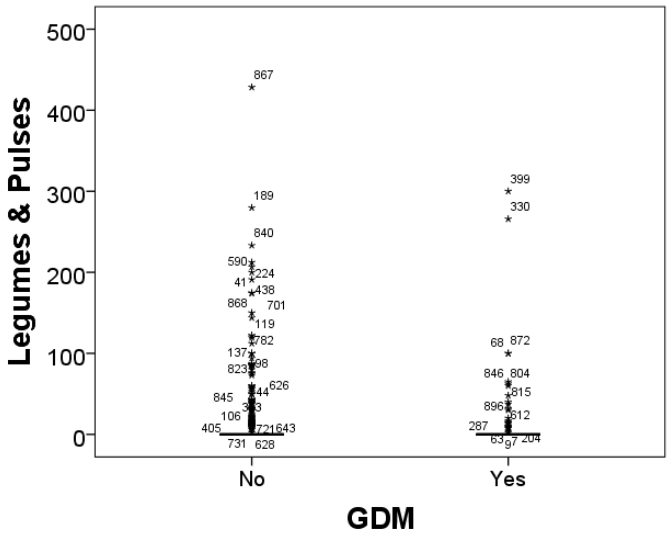
Unknown match factors are due to manual integration of compounds.

Compound class only listed for metabolites detected with a match factor > 60 %.

Appendix 10. Box plots of food groups in the canonical correlation roots







References

Adams, S. H. (2011). Emerging perspectives on essential amino acid metabolism in obesity and the insulin-resistant state. *Advances in Nutrition: An International Review Journal, 2*(6), 445–456. http://doi.org/10.3945/an.111.000737

Aggio, R., Villas-Bôas, S. G., & Ruggiero, K. (2011). Metab: An R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics, 27*(16), 2316–2318. http://doi.org/10.1093/bioinformatics/btr379

Ajala, O., English, P., & Pinkney, J. (2013). Systematic review and meta-analysis of different dietary approaches to the management of type 2 diabetes. *The American journal of clinical nutrition, 97*(3), 505-516.

Akhlaghi, F., Bagheri, S. M., & Rajabi, O. (2012). A comparative study of relationship between micronutrients and gestational diabetes. *ISRN Obstetrics and Gynecology, 2012*, 1–4. https://doi.org/10.5402/2012/470419

Alberti, K. G., & Zimmet, P. Z. (1998). Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. Provisional report of a WHO Consultation. *Diabetic Medicine, 15*(7), 539–553. http://doi.org/10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S

American Diabetes Association, Bantle, J. P., Wylie-Rosett, J., Albright, A. L., Apovian, C. M., Clark, N. G., … Wheeler, M. L. (2008). Nutrition recommendations and interventions for diabetes: A position statement of the American Diabetes Association. *Diabetes Care, 31*(Supplement 1), S61–S78. https://doi.org/10.2337/dc08-S061

Anandakumar, C., Wong, Y. C., Annapoorna, V., Arulkumaran, S., Chia, D., Bongso, A., & Ratnam, S. S. (1992). Amniocentesis and its complications. *The Australian and New Zealand Journal of Obstetrics and Gynaecology, 32*(2), 97–99. https://doi.org/10.1111/j.1479-828X.1992.tb01916.x

Arnold, D. L., Enquobahrie, D. A., Qiu, C., Huang, J., Grote, N., VanderStoep, A., & Williams, M. A. (2015). Early pregnancy maternal vitamin D concentrations and risk of gestational diabetes mellitus. *Paediatric and Perinatal Epidemiology, 29*(3), 200–210. https://doi.org/10.1111/ppe.12182

Azaïs-Braesco, V., & Pascal, G. (2000). Vitamin A in pregnancy: Requirements and safety limits. *American Society for Clinical Nutrition, 71*(5), 1325s–1333s.

Backwell, L., Pickering, R., Brothwell, D., Berger, L., Witcomb, M., Martill, D., … Wilson, A. (2009). Probable human hair found in a fossil hyaena coprolite from Gladysvale cave, South Africa. *Journal of Archaeological Science, 36*(6), 1269–1276. http://doi.org/10.1016/j.jas.2009.01.023

Baker, A. M., Haeri, S., Camargo, C. A., Stuebe, A. M., & Boggess, K. A. (2012). First-trimester maternal vitamin D status and risk for gestational diabetes (GDM): A nested case-control study. *Diabetes/Metabolism Research and Reviews, 28*(2), 164–168. https://doi.org/10.1002/dmrr.1282

Bando, K., Kawahara, R., Kunimatsu, T., Sakai, J., Kimura, J., Funabashi, H., … Fukusaki, E. (2010). Influences of biofluid sample collection and handling procedures on GC–MS based metabolomic studies. *Journal of Bioscience and Bioengineering, 110*(4), 491–499. http://doi.org/10.1016/j.jbiosc.2010.04.010

Bao, W., Bowers, K., Tobias, D. K., Hu, F. B., & Zhang, C. (2013). Prepregnancy dietary protein intake, major dietary protein sources, and the risk of gestational diabetes mellitus: A prospective cohort study. *Diabetes Care*, *36*(7), 2001–2008. https://doi.org/10.2337/dc12-2018

Bao, W., Bowers, K., Tobias, D. K., Olsen, S. F., Chavarro, J., Vaag, A., … Zhang, C. (2014). Prepregnancy low-carbohydrate dietary pattern and risk of gestational diabetes mellitus: A prospective cohort study. *The American Journal of Clinical Nutrition*, *99*(6), 1378–1384. https://doi.org/10.3945/ajcn.113.082966

Bao, W., Tobias, D. K., Hu, F. B., Chavarro, J. E., & Zhang, C. (2016). Pre-pregnancy potato consumption and risk of gestational diabetes mellitus: Prospective cohort study. *BMJ*, h6898. https://doi.org/10.1136/bmj.h6898

Bao, W., Tobias, D. K., Olsen, S. F., & Zhang, C. (2014). Pre-pregnancy fried food consumption and the risk of gestational diabetes mellitus: A prospective cohort study. *Diabetologia*, *57*(12), 2485–2491. https://doi.org/10.1007/s00125-014-3382-x

Barbieri, M., Rosaria Rizzo, M., Manzella, D., & Paolisso, G. (2001). Age-related insulin resistance: Is it an obligatory finding? The lesson from healthy centenarians. *Diabetes/Metabolism Research and Reviews*, *17*(1), 19-26.

Bartáková, V., Kuricová, K., Zlámal, F., Bělobrádková, J., & Kaňková, K. (2016). Differences in food intake and genetic variability in taste receptors between Czech pregnant women with and without gestational diabetes mellitus. *European Journal of Nutrition*. https://doi.org/10.1007/s00394-016-1334-6

Bartáková, V., Pleskačová, A., Kuricová, K., Pácal, L., Dvořáková, V., Bělobrádková, J., … Kaňková, K. (2016). Dysfunctional protection against advanced glycation due to thiamine metabolism abnormalities in gestational diabetes. *Glycoconjugate Journal*, *33*(4), 591–598. https://doi.org/10.1007/s10719-016-9688-9

Bastard, J.-P., Maachi, M., Lagathu, C., Kim, M. J., Caron, M., Vidal, H., … Feve, B. (2006). Recent advances in the relationship between obesity, inflammation, and insulin resistance. *European Cytokine Network*, *17*(1), 4–12.

Behboudi-Gandevani, S., Safary, K., Moghaddam-Banaem, L., Lamyian, M., Goshtasbi, A., & Alian-Moghaddam, N. (2013). The relationship between maternal serum iron and zinc levels and their nutritional intakes in early pregnancy with gestational diabetes. *Biological Trace Element Research*, *154*(1), 7–13. https://doi.org/10.1007/s12011-013-9703-y

Bellamy, L., Casas, J.-P., Hingorani, A. D., & Williams, D. (2009a). Type 2 diabetes mellitus after gestational diabetes: A systematic review and meta-analysis. *The Lancet*, *373*(9677), 1773–1779. https://doi.org/10.1016/S0140-6736(09)60731-5

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300. http://www.jstor.org/stable/2346101

Bistritzer, T., Roeder, L., Hanukoglu, L., & Levin, P. (1991). Non-insulin dependent diabetic patients have increased glucose uptake in red blood cells. *Hormone and Metabolic Research*, *23*(02), 70–73. http://doi.org/10.1055/s-2007-1003616

Bloom, D. E., Cafiero, E. T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A. B., Gaziano, T., Mowafi, M., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A.Z., & Weinstein, C. (2011). *The global economic burden of non-*

communicable diseases. Geneva, Switzerland: World Economic Forum.

Boney, C. M., Verma, A., Tucker, R., & Vohr, B. R. (2005). Metabolic syndrome in childhood: Association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics*, *115*(3), e290-296. https://doi.org/10.1542/peds.2004-1808

Bothwell, T. H. (2000). Iron requirements in pregnancy and strategies to meet them. *The American Journal of Clinical Nutrition*, *72*(1), 257s–264s.

Bouatra, S., Aziat, F., Mandal, R., Guo, A. C., Wilson, M. R., Knox, C., … Cruz, N. (2013). The human urine metabolome. *PLoS ONE*, *8*(9), e73076. http://doi.org/10.1371/journal.pone.0073076

Bouchard-Mercier, A., Rudkowska, I., Lemieux, S., Couture, P., & Vohl, M.-C. (2013). The metabolic signature associated with the Western dietary pattern: A cross-sectional study. *Nutrition Journal*, *12*, 158. https://doi.org/10.1186/1475-2891-12-158

Bowers, K. A., Olsen, S. F., Bao, W., Halldorsson, T. I., Strom, M., & Zhang, C. (2016). Plasma concentrations of ferritin in early pregnancy are associated with risk of gestational diabetes mellitus in women in the Danish National Birth Cohort. *Journal of Nutrition*, *146*(9), 1756–1761. https://doi.org/10.3945/jn.115.227793

Bowers, K., Tobias, D. K., Yeung, E., Hu, F. B., & Zhang, C. (2012). A prospective study of prepregnancy dietary fat intake and risk of gestational diabetes. *American Journal of Clinical Nutrition*, *95*(2), 446–453. https://doi.org/10.3945/ajcn.111.026294

Bowers, K., Yeung, E., Williams, M. A., Qi, L., Tobias, D. K., Hu, F. B., & Zhang, C. (2011). A prospective study of prepregnancy dietary iron intake and risk for gestational diabetes mellitus. *Diabetes Care*, *34*(7), 1557–1563. https://doi.org/10.2337/dc11-0134

Boyko, E. J., Fujimoto, W. Y., Leonetti, D. L., & Newell-Morris, L. (2000). Visceral adiposity and risk of type 2 diabetes: A prospective study among Japanese Americans. *Diabetes Care*, *23*(4), 465–471.

Boyle, V. T., Thorstensen, E. B., Mourath, D., Jones, M. B., McCowan, L. M., Kenny, L. C., & Baker, P. N. (2016). The relationship between 25-hydroxyvitamin D concentration in early pregnancy and pregnancy outcomes in a large, prospective cohort. *British Journal of Nutrition*, *116*(8), 1409–1415. https://doi.org/10.1017/S0007114516003202

Brantsaeter, A. L., Haugen, M., Samuelsen, S. O., Torjusen, H., Trogstad, L., Alexander, J., … Meltzer, H. M. (2009). A dietary pattern characterized by high intake of vegetables, fruits, and vegetable oils is associated with reduced risk of preeclampsia in nulliparous pregnant Norwegian women. *Journal of Nutrition*, *139*(6), 1162–1168. http://doi.org/10.3945/jn.109.104968

Bray, G. A. (2004). Medical consequences of obesity. *The Journal of Clinical Endocrinology & Metabolism*, *89*(6), 2583–2589.

Brown, L. (2011). Nutrition requirements during pregnancy. In J. Sharlin & S. Edelstein (Eds.), *Essentials of life cycle nutrition* (pp. 1–24). Sudbury, MA: Jones and Bartlett.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. http://doi.org/10.1038/nrn3475

Camprubi Robles, M., Campoy, C., Garcia Fernandez, L., Lopez-Pedrosa, J. M., Rueda, R., & Martin, M. J. (2015). Maternal diabetes and cognitive performance in the offspring: A

systematic review and meta-analysis. *PloS One*, *10*(11), e0142583. https://doi.org/10.1371/journal.pone.0142583

Casey, B. M., Lucas, M. J., Mcintire, D. D., & Leveno, K. J. (1997). Pregnancy outcomes in women with gestational diabetes compared with the general obstetric population. *Obstetrics and Gynecology*, *90*(6), 869–873.

Catalano, P. M., Ishizuka, T., & Friedman, J. E. (1998). Glucose metabolism in pregnancy. In R. M. Cowett (Ed.), *Principles of Perinatal-Neonatal Metabolism* (pp. 183–206). New York, NY: Springer-Verlag.

Chan, K., Chan, B., Lam, K., Tam, S., & Lao, T. (2009). Iron supplement in pregnancy and development of gestational diabetes-a randomised placebo-controlled trial. *BJOG: An International Journal of Obstetrics & Gynaecology*, *116*(6), 789–798. https://doi.org/10.1111/j.1471-0528.2008.02014.x

Chen, X., Scholl, T. O., Leskiw, M., Savaille, J., & Stein, T. P. (2010). Differences in maternal circulating fatty acid composition and dietary fat intake in women with gestational diabetes mellitus or mild gestational hyperglycemia. *Diabetes Care*, *33*(9), 2049–2054. https://doi.org/10.2337/dc10-0693

Chen, Y., Quick, W. W., Yang, W., Zhang, Y., Baldwin, A., Moran, J., … Dall, T. M. (2009). Cost of gestational diabetes mellitus in the United States in 2007. *Population Health Management*, *12*(3), 165–174. http://doi.org/10.1089/pop.2009.12303

Chong, Y., Cai, S., Lin, H., Soh, S., Lee, Y., Leow, M., … Kwek, K. (2014). Ethnic differences translate to inadequacy of high-risk screening for gestational diabetes mellitus in an Asian population: A cohort study. *BMC Pregnancy and Childbirth*, *14*, 345. https://doi.org/10.1186/1471-2393-14-345

Chu, S. Y., Callaghan, W. M., Kim, S. Y., Schmid, C. H., Lau, J., England, L. J., & Dietz, P. M. (2007). Maternal obesity and risk of gestational diabetes mellitus. *Diabetes Care*, *30*(8), 2070–2076.

Conway, J. M., Ingwersen, L. A., Vinyard, B. T., & Moshfegh, A. J. (2003). Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women. *The American Journal of Clinical Nutrition*, *77*(5), 1171–1178.

Correa, P. J., Vargas, J. F., Sen, S., & Illanes, S. E. (2014). Prediction of gestational diabetes early in pregnancy: Targeting the long-term complications. *Gynecologic and Obstetric Investigation*, *77*(3), 145–149. http://doi.org/10.1159/000357616

Covaci, A., Tutudaki, M., Tsatsakis, A. M., & Schepens, P. (2002). Hair analysis: Another approach for the assessment of human exposure to selected persistent organochlorine pollutants. *Chemosphere*, *46*(3), 413–418. http://doi.org/10.1016/S0045-6535(01)00065-0

Crozier, S. R., Robinson, S. M., Borland, S. E., Inskip, H. M., & SWS Study Group. (2006). Dietary patterns in the Southampton Women's Survey. *European journal of clinical nutrition*, *60*(12), 1391.

Crozier, S. R., Robinson, S. M., Godfrey, K. M., Cooper, C., & Inskip, H. M. (2009). Women's dietary patterns change little from before to during pregnancy. *The Journal of Nutrition*, *139*(10), 1956-1963.

Cucó, G., Fernandez-Ballart, J., Sala, J., Viladrich, C., Iranzo, R., Vila, J., & Arija, V. (2006).

Dietary patterns and associated lifestyles in preconception, pregnancy and postpartum. *European Journal of Clinical Nutrition*, *60*(3), 364-371.

D'Anna, R., Di Benedetto, V., Rizzo, P., Raffone, E., Interdonato, M. L., Corrado, F., & Di Benedetto, A. (2012). Myo-inositol may prevent gestational diabetes in PCOS women. *Gynecological Endocrinology*, *28*(6), 440–442. https://doi.org/10.3109/09513590.2011.633665

Darling, A., Mitchell, A., & Werler, M. (2016). Preconceptional iron intake and gestational diabetes mellitus. *International Journal of Environmental Research and Public Health*, *13*(6), 525. https://doi.org/10.3390/ijerph13060525

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., … Turnbaugh, P. J. (2013). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559–563. http://doi.org/10.1038/nature12820

De Filippis, F., Pellegrini, N., Vannini, L., Jeffery, I. B., La Storia, A., Laghi, L., … Ercolini, D. (2016). High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut*, *65*(11), 1812–1821. https://doi.org/10.1136/gutjnl-2015-309957

de Seymour, J. V, Conlon, C. A., Sulek, K., Villas Bôas, S. G., McCowan, L. M., Kenny, L. C., & Baker, P. N. (2014). Early pregnancy metabolite profiling discovers a potential biomarker for the subsequent development of gestational diabetes mellitus. *Acta Diabetologica*, *51*(5), 887–890. https://doi.org/10.1007/s00592-014-0626-7

Deeming, S. B., & Weber, C. W. (1978). Hair analysis of trace minerals in human subjects as influenced by age, sex, and contraceptive drugs. *The American Journal of Clinical Nutrition*, *31*(7), 1175–1180.

Delange, F. (2007). Iodine requirements during pregnancy, lactation and the neonatal period and indicators of optimal iodine nutrition. *Public Health Nutrition*, *10*(12A), 1571–1580.

Derbent, A. U., Simavli, S. A., Kaygusuz, I., Gumus, I. I., Yılmaz, S., Yıldırım, M., & Uysal, S. (2013). Serum hepcidin is associated with parameters of glucose metabolism in women with gestational diabetes mellitus. *The Journal of Maternal-Fetal & Neonatal Medicine*, *26*(11), 1112–1115. https://doi.org/10.3109/14767058.2013.770462

Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, *26*(1), 51–78. http://doi.org/10.1002/mas.20108

Deurenberg, P., Deurenberg-Yap, M., & Guricci, S. (2002). Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship. *Obesity Reviews: An Official Journal of the International Association for the Study of Obesity*, *3*(3), 141–146.

Deurenberg-Yap, M., Schmidt, G., van Staveren, W. A., & Deurenberg, P. (2000). The paradox of low body mass index and high body fat percentage among Chinese, Malays and Indians in Singapore. *International Journal of Obesity and Related Metabolic Disorders: Journal of the International Association for the Study of Obesity*, *24*(8), 1011–1017.

Diaz, S. O., Barros, A. S., Goodfellow, B. J., Duarte, I. F., Galhano, E., Pita, C., … Gil, A. M. (2013). Second trimester maternal urine for the diagnosis of trisomy 21 and prediction of poor pregnancy outcomes. *Journal of Proteome Research*, *12*(6), 2946–2957. https://doi.org/10.1021/pr4002355

Diaz, S. O., Pinto, J., Graça, G., Duarte, I. F., Barros, A. S., Galhano, E., … Gil, A. M. (2011). Metabolic biomarkers of prenatal disorders: An exploratory NMR metabonomics study of second trimester maternal urine and blood plasma. *Journal of Proteome Research*, *10*(8), 3732–3742. https://doi.org/10.1021/pr200352m

Dodds, L., Woolcott, C. G., Weiler, H., Spencer, A., Forest, J.-C., Armson, B. A., & Giguère, Y. (2016). Vitamin D status and gestational diabetes: Effect of smoking status during pregnancy. *Paediatric and Perinatal Epidemiology*, *30*(3), 229–237. https://doi.org/10.1111/ppe.12278

Domaracki, P., Sadlecki, P., Odrowaz-Sypniewska, G., Dzikowska, E., Walentowicz, P., Siodmiak, J., … Walentowicz-Sadlecka, M. (2016). Serum 25(OH) Vitamin D levels in Polish women during pregnancies complicated by hypertensive disorders and gestational diabetes. *International Journal of Molecular Sciences*, *17*(10), 1574. https://doi.org/10.3390/ijms17101574

Dudzik, D., Zorawski, M., Skotnicki, M., Zarzycki, W., Kozlowska, G., Bibik-Malinowska, K., ... & Ramos, M. P. (2014). Metabolic fingerprint of gestational diabetes mellitus. *Journal of proteomics*, *103*, 57-71.

Dudzik, D., Zorawski, M., Skotnicki, M., Zarzycki, W., García, A., Angulo, S., ... & Ramos, M. P. (2017). GC–MS based Gestational Diabetes Mellitus longitudinal study: Identification of 2-and 3-hydroxybutyrate as potential prognostic biomarkers. *Journal of Pharmaceutical and Biomedical Analysis*, *144*, 90-98

Duncan, B. B., Inês Schmidt, M., Pankow, J. S., Ballantyne, C. M., Couper, D., Vigo, A., … Heiss, G. (2003). Low-grade systemic inflammation and the development of type 2 diabetes. *Diabetes*, *52*(7), 1799–1805.

Dunn, W., Broadhurst, D., Atherton, H., Goodacre, R., & Griffin, J. (2011). Systems level studies of mammalian metabolomes: The roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews*, *40*, 387–426.

Edmands, W. M., Ferrari, P., Rothwell, J. A., Rinaldi, S., Slimani, N., Barupal, D. K., … Scalbert, A. (2015). Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. *American Journal of Clinical Nutrition*, *102*(4), 905–913. https://doi.org/10.3945/ajcn.114.101881

Ehrenberg, H. M., Mercer, B. M., & Catalano, P. M. (2004). The influence of obesity and diabetes on the prevalence of macrosomia. *American Journal of Obstetrics and Gynecology*, *191*(3), 964–968. https://doi.org/10.1016/j.ajog.2004.05.052

Emwas, A.-H. M. (2015). The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods in Molecular Biology*, *1277*, 161–193. https://doi.org/10.1007/978-1-4939-2377-9_13

Englund-Ögge, L., Brantsæter, A. L., Sengpiel, V., Haugen, M., Birgisdottir, B. E., Myhre, R., … Jacobsson, B. (2014). Maternal dietary patterns and preterm delivery: Results from large prospective cohort study. *BMJ*, *348*, g1446.

Enquobahrie, D. A., Denis, M., Tadesse, M. G., Gelaye, B., Ressom, H. W., & Williams, M. A. (2015). Maternal early pregnancy serum metabolites and risk of gestational diabetes mellitus. *The Journal of Clinical Endocrinology & Metabolism*, *100*(11), 4348–4356. http://doi.org/10.1210/jc.2015-2862

Fanos, V., Atzori, L., Makarenko, K., Melis, G. B., & Ferrazzi, E. (2013). Metabolomics application in maternal-fetal medicine. *BioMed research international*, June, 2013.

Farland, L. V., Rifas-Shiman, S. L., & Gillman, M. W. (2015). Early pregnancy cravings, dietary intake, and development of abnormal glucose tolerance. *Journal of the Academy of Nutrition and Dietetics*, *115*(12), 1958–1964.e1. https://doi.org/10.1016/j.jand.2015.04.018

Farrant, H. J., Krishnaveni, G. V, Hill, J. C., Boucher, B. J., Fisher, D. J., Noonan, K., … Fall, C. H. (2009). Vitamin D insufficiency is common in Indian mothers but is not associated with gestational diabetes or variation in newborn size. *European Journal of Clinical Nutrition*, *63*(5), 646–652. https://doi.org/10.1038/ejcn.2008.14

Fetissov, S. O. (2016). Role of the gut microbiota in host appetite control: Bacterial growth to animal feeding behaviour. *Nature Reviews Endocrinology*, *13*(1), 11–25. http://doi.org/10.1038/nrendo.2016.150

Fetita, L.-S., Sobngwi, E., Serradas, P., Calvo, F., & Gautier, J.-F. (2006). Consequences of fetal exposure to maternal diabetes in offspring. *The Journal of Clinical Endocrinology & Metabolism*, *91*(10), 3718–3724. https://doi.org/10.1210/jc.2006-0624

Fiehn, O., Garvey, W. T., Newman, J. W., Lok, K. H., Hoppel, C. L., Adams, S. H., … Ricquier, D. (2010). Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PLoS ONE*, *5*(12), e15234. http://doi.org/10.1371/journal.pone.0015234

Floegel, A., von Ruesten, A., Drogan, D., Schulze, M. B., Prehn, C., Adamski, J., … Boeing, H. (2013). Variation of serum metabolites related to habitual diet: A targeted metabolomic approach in EPIC-Potsdam. *European Journal of Clinical Nutrition*, *67*(10), 1100–1108. https://doi.org/10.1038/ejcn.2013.147

Flood-Nichols, S. K., Tinnemore, D., Huang, R. R., Napolitano, P. G., & Ippolito, D. L. (2015). Vitamin D deficiency in early pregnancy. *PLOS ONE*, *10*(4), e0123763. https://doi.org/10.1371/journal.pone.0123763

Fontana, L., Eagon, C. J., Trujillo, M. E., Scherer, P. E., & Klein, S. (2007). Visceral fat adipokine secretion is associated with systemic inflammation in obese humans. *Diabetes*, *56*(4), 1010–1013.

Franco, R. S. (2012). Measurement of red cell lifespan and aging. *Transfus Med Hemother*, *39*(5), 302-307

Freitas Lima, L., Braga, V., do Socorro de França Silva, M., Cruz, J., Sousa Santos, S., de Oliveira Monteiro, M., & Balarini, C. (2015). Adipokines, diabetes and atherosclerosis: An inflammatory association. *Frontiers in Physiology*, *6*, 304.

Furukawa, S., Fujita, T., Shimabukuro, M., Iwaki, M., Yamada, Y., Nakajima, Y., … Goldstein, J. (2004). Increased oxidative stress in obesity and its impact on metabolic syndrome. *The Journal of Clinical Investigation*, *114*(12), 1752–1761. http://doi.org/10.1172/JCI21625

Gall, W. E., Beebe, K., Lawton, K. A., Adam, K.-P., Mitchell, M. W., Nakhle, P. J., … Wiener, M. (2010). α-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS ONE*, *5*(5), e10883. http://doi.org/10.1371/journal.pone.0010883

Garcia-Aloy, M., Llorach, R., Urpi-Sarda, M., Jáuregui, O., Corella, D., Ruiz-Canela, M., … Andres-Lacueva, C. (2015). A metabolomics-driven approach to predict cocoa product consumption by designing a multimetabolite biomarker model in free-living subjects from the PREDIMED study. *Molecular Nutrition & Food Research*, *59*(2), 212–220.

https://doi.org/10.1002/mnfr.201400434

Garcia-Aloy, M., Llorach, R., Urpi-Sarda, M., Tulipani, S., Estruch, R., Martínez-González, M. A., … Andres-Lacueva, C. (2014). Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: The PREDIMED Study. *Journal of Proteome Research*, *13*(7), 3476–3483. https://doi.org/10.1021/pr500425r

Garcia-Aloy, M., Llorach, R., Urpi-Sarda, M., Tulipani, S., Salas-Salvadó, J., Martínez-González, M. A., … Andres-Lacueva, C. (2015). Nutrimetabolomics fingerprinting to identify biomarkers of bread exposure in a free-living population from the PREDIMED study cohort. *Metabolomics*, *11*(1), 155–165. https://doi.org/10.1007/s11306-014-0682-6

Garcia-Perez, I., Posma, J. M., Gibson, R., Chambers, E. S., Hansen, T. H., Vestergaard, H., … Frost, G. (2017). Objective assessment of dietary patterns by use of metabolic phenotyping: A randomised, controlled, crossover trial. *The Lancet Diabetes & Endocrinology*, *5*(3), 184–195. https://doi.org/10.1016/S2213-8587(16)30419-3

Gibbons, H., McNulty, B. A., Nugent, A. P., Walton, J., Flynn, A., Gibney, M. J., & Brennan, L. (2015). A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake. *American Journal of Clinical Nutrition*, *101*(3), 471–477. https://doi.org/10.3945/ajcn.114.095604

Gibney, M. J., Walsh, M., Brennan, L., Roche, H. M., German, B., & van Ommen, B. (2005). Metabolomics in human nutrition: Opportunities and challenges. *The American Journal of Clinical Nutrition*, *82*(3), 497–503.

Goker Tasdemir, U., Tasdemir, N., Kilic, S., Abali, R., Celik, C., & Gulerman, H. C. (2015). Alterations of ionized and total magnesium levels in pregnant women with gestational diabetes mellitus. *Gynecologic and Obstetric Investigation*, *79*(1), 19–24. https://doi.org/10.1159/000365813

Gonzalez-Covarrubias, V., Dane, A., Hankemeier, T., & Vreeken, R. J. (2013). The influence of citrate, EDTA, and heparin anticoagulants to human plasma LC–MS lipidomic profiling. *Metabolomics*, *9*(2), 337–348. http://doi.org/10.1007/s11306-012-0450-4

Gowda, G. A., Zhang, S., Gu, H., Asiago, V., Shanaiah, N., & Raftery, D. (2008). Metabolomics-based methods for early disease diagnostics. *Expert Review of Molecular Diagnostics*, *8*(5), 617–633. https://doi.org/10.1586/14737159.8.5.617

Graça, G., Duarte, I. F., Barros, A. S., Goodfellow, B. J., Diaz, S. O., Pinto, J., … Gil, A. M. (2010). Impact of prenatal disorders on the metabolic profile of second trimester amniotic fluid: A nuclear magnetic resonance metabonomic study. *Journal of Proteome Research*, *9*(11), 6016–6024. https://doi.org/10.1021/pr100815q

Graça, G., Goodfellow, B. J., Barros, A. S., Diaz, S., Duarte, I. F., Spagou, K., … Gil, A. M. (2012). UPLC-MS metabolic profiling of second trimester amniotic fluid and maternal urine and comparison with NMR spectral profiling for the identification of pregnancy disorder biomarkers. *Molecular bioSystems*, *8*(4), 1243–1254. https://doi.org/10.1039/c2mb05424h

Graham, K., Koren, G., Klein, J., Schneiderman, J., & Greenwald, M., (1989). Determination of gestational cocaine exposure by hair analysis. *JAMA: The Journal of the American Medical Association*, *262*(23), 3328-3330. http://doi.org/10.1001/jama.1989.03430230113036

Gresham, E., Collins, C. E., Mishra, G. D., Byles, J. E., & Hure, A. J. (2016). Diet quality before or during pregnancy and the relationship with pregnancy and birth outcomes: The Australian Longitudinal Study on Women's Health. *Public Health Nutrition*, *19*(16), 2975–2983. https://doi.org/10.1017/S1368980016001245

Grissaa, O., Atègbod, J.-M., Yessoufoud, A., Tabkaa, Z., Miledb, A., Jerbic, M., … Khand, N. . (2007). Antioxidant status and circulating lipids are altered in human gestational diabetes and macrosomia. *Translational Research*, *150*(3), 164–171.

Haidari, F., Jalali, M.-T., Shahbazian, N., Haghighizadeh, M.-H., & Azadegan, E. (2016). Comparison of serum levels of vitamin D and inflammatory markers between women with gestational diabetes mellitus and healthy pregnant control. *Journal of Family & Reproductive Health*, *10*(1), 1–8.

Hajjawi, O. S. (2013). Glucose transport in human red blood cells. *American Journal of Biomedical and Life Sciences*, *1*(3), 44–52. http://doi.org/10.11648/j.ajbls.20130103.12

Hamdan, H. Z., Elbashir, L. M., Hamdan, S. Z., Elhassan, E. M., & Adam, I. (2014). Zinc and selenium levels in women with gestational diabetes mellitus at Medani Hospital, Sudan. *Journal of Obstetrics and Gynaecology*, *34*(7), 567–570. https://doi.org/10.3109/01443615.2014.919576

Hanley, A. J., Wagenknecht, L. E., Haffner, J. M., Norris, M., Bryer-Ash, Y. I., Chen, A. M., … Bergman, S. M. (2009). Insulin resistance, beta cell dysfunction and visceral adiposity as predictors of incident diabetes: The insulin resistance atherosclerosis study (IRAS) family study. *Diabetologia*, *52*(10), 2079–2086.

Harkey, M. R. (1993). Anatomy and physiology of hair. *Forensic Science International*, *63*(1), 9–18. http://doi.org/10.1016/0379-0738(93)90255-9

He, J.-R., Yuan, M.-Y., Chen, N.-N., Lu, J.-H., Hu, C.-Y., Mai, W.-B., … Qiu, X. (2015). Maternal dietary patterns and gestational diabetes mellitus: A large prospective cohort study in China. *British Journal of Nutrition*, *113*(8), 1292–1300. https://doi.org/10.1017/S0007114515000707

He, X., de Seymour, J. V., Sulek, K., Qi, H., Zhang, H., Han, T.-L., … Baker, P. N. (2016). Maternal hair metabolome analysis identifies a potential marker of lipid peroxidation in gestational diabetes mellitus. *Acta Diabetologica*, *53*(1), 119–122. https://doi.org/10.1007/s00592-015-0737-9

Hedderson, M. M., Ferrara, A., & Sacks, D. A. (2003). Gestational diabetes mellitus and lesser degrees of pregnancy hyperglycemia: Association with increased risk of spontaneous preterm birth. *Obstetrics and Gynecology*, *102*(4), 850–856.

Hedderson, M. M., Gunderson, E. P., & Ferrara, A. (2010). Gestational weight gain and risk of gestational diabetes mellitus. *Obstetrics and Gynecology*, *115*(3), 597–604. https://doi.org/10.1097/AOG.0b013e3181cfce4f

Heinzmann, S. S., Brown, I. J., Chan, Q., Bictash, M., Dumas, M.-E., Kochhar, S., … Nicholson, J. K. (2010). Metabolic profiling strategy for discovery of nutritional biomarkers: Proline betaine as a marker of citrus consumption. *American Journal of Clinical Nutrition*, *92*(2), 436–443. https://doi.org/10.3945/ajcn.2010.29672

Hekmat, K., Bagheri, R., Abedi, P., & Tabesh, H. (2014). The relationship of fat soluble antioxidants with gestational diabetes in Iran: A case-control study. *The Journal of Maternal-Fetal & Neonatal Medicine*, *27*(16), 1676–1679. https://doi.org/10.3109/14767058.2014.892921

Helin, A., Kinnunen, T. I., Raitanen, J., Ahonen, S., Virtanen, S. M., & Luoto, R. (2012). Iron intake, haemoglobin and risk of gestational diabetes: A prospective cohort study. *BMJ Open*, *2*(5), e001730–e001730. https://doi.org/10.1136/bmjopen-2012-001730

Hex, N., Bartlett, C., Wright, D., Taylor, M., & Varley, D. (2012). Estimating the current and future costs of type 1 and type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabetic Medicine*, *29*(7), 855–862. http://doi.org/10.1111/j.1464-5491.2012.03698.x

Heymann, D. L., & American Public Health Association. (2008). *Control of communicable diseases manual*. Washington, DC: American Public Health Association.

Hodge, A. M., English, D. R., O'Dea, K., & Giles, G. G. (2004). Glycemic index and dietary fiber and the risk of type 2 diabetes. *Diabetes Care*, *27*(11), 2701–2706. http://doi.org/10.2337/diacare.27.11.2701

Hoffmann, K., Schulze, M. B., Schienkiewitz, A., Nöthlings, U., & Boeing, H. (2004). Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *American journal of epidemiology*, *159*(10), 935-944.

Højlund, K., Wrzesinski, K., Larsen, P., Fey, S., Roepstorff, P., Handberg, A., … Beck-Nielsen, H. (2003). Proteome analysis reveals phosphorylation of ATP synthase beta -subunit in human skeletal muscle and proteins with potential roles in type 2 diabetes. *J Biol Chem.*, *278*(12), 10436–10442.

Horgan, R. P., Clancy, O. H., Myers, J. E., & Baker, P. N. (2009). An overview of proteomic and metabolomic technologies and their application to pregnancy research. *BJOG: An International Journal of Obstetrics & Gynaecology*, *116*(2), 173-181.

Horgan, R. P., Broadhurst, D. I., Walsh, S. K., Dunn, W. B., Brown, M., Roberts, C. T., … Kenny, L. C. (2011). Metabolic profiling uncovers a phenotypic signature of small for gestational age in early pregnancy. *Journal of Proteome Research*, *10*(8), 3660–3673. https://doi.org/10.1021/pr2002897

Horvath, A. L. (2009). Solubility of structurally complicated materials: 3. Hair. *The Scientific World JOURNAL*, *9*, 255–271. http://doi.org/10.1100/tsw.2009.27

Hotamisligil, G. S. (2006). Inflammation and metabolic disorders. *Nature*, *444*(7121), 860-867.

Hu, F. B., Manson, J. E., Stampfer, M. J., Colditz, G., Liu, S., Solomon, C. G., & Willett, W. C. (2001). Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *New England journal of medicine*, *345*(11), 790-797.

Hu, F. B. (2002). Dietary pattern analysis: A new direction in nutritional epidemiology. *Current Opinion in Lipidology*, *13*(1), 3–9.

Hu, F. B. (2011). Globalization of diabetes: The role of diet, lifestyle, and genes. *Diabetes Care*, *34*(6), 1249–1257. https://doi.org/10.2337/dc11-0442

Huynh, J., Xiong, G., & Bentley-Lewis, R. (2014). A systematic review of metabolite profiling in gestational diabetes mellitus. *Diabetologia*, *57*(12), 2453–2464. https://doi.org/10.1007/s00125-014-3371-0

Inoue, K., Tanada, C., Hosoya, T., Yoshida, S., Akiba, T., Min, J. Z., … Toyo'oka, T. (2016). Principal component analysis of molecularly based signals from infant formula contaminations using LC-MS and NMR in foodomics. *J Sci Food Agric*, *96*(11) 3876–3881. https://doi.org/10.1002/jsfa.7584

International Diabetes Federation. (2015). *IDF Diabetes Atlas* (7th ed). Retrieved from www.diabetesatlas.org

Ioannidis, J. P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *JAMA, 294*(2), 218-228. http://doi.org/10.1001/jama.294.2.218

Izadi, V., Tehrani, H., Haghighatdoost, F., Dehghan, A., Surkan, P. J., & Azadbakht, L. (2016). Adherence to the DASH and Mediterranean diets is associated with decreased risk for gestational diabetes mellitus. *Nutrition, 32*(10), 1092–1096. https://doi.org/10.1016/j.nut.2016.03.006

Jacobs, D. R., & Tapsell, L. C. (2007). Food, not nutrients, is the fundamental unit in nutrition. *Nutrition Reviews, 65*(10), 439–450.

Karamanos, B., Thanopoulou, A., Anastasiou, E., Assaad-Khalil, S., Albache, N., Bachaoui, M., … Savona-Ventura, C. (2014). Relation of the Mediterranean diet with the incidence of gestational diabetes. *European Journal of Clinical Nutrition, 68*(1), 8–13. https://doi.org/10.1038/ejcn.2013.177

Kelley, D., He, J., Menshikova, E., & Ritov, V. (2002). Dysfunction of mitochondria in human skeletal muscle in type 2 diabetes. *Diabetes, 51*(10), 2944–2950.

Kenny, L. C., Broadhurst, D. I., Dunn, W., Brown, M., North, R. A., McCowan, L., … Screening for Pregnancy Endpoints Consortium. (2010). Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. *Hypertension, 56*(4), 741–749. https://doi.org/10.1161/HYPERTENSIONAHA.110.157297

Kershaw, E., & Flier, J. (2004). Adipose tissue as an endocrine organ. *J Clin Endocrinol Metab, 89*(6), 2548–2556.

Khambalia, A. Z., Aimone, A., Nagubandi, P., Roberts, C. L., McElduff, A., Morris, J. M., … Nassar, N. (2015). High maternal iron status, dietary iron intake and iron supplement use in pregnancy and risk of gestational diabetes mellitus: A prospective study and systematic review. *Diabetic Medicine, 33*(9), 1211-1221. https://doi.org/10.1111/dme.13056

Kilinc, M., Guven, M. A., Ezer, M., Ertas, I. E., & Coskun, A. (2008). Evaluation of serum selenium levels in Turkish women with gestational diabetes mellitus, glucose intolerants, and normal controls. *Biological Trace Element Research, 123*(1–3), 35–40. https://doi.org/10.1007/s12011-007-8087-2

Kolu, P., Raitanen, J., Rissanen, P., & Luoto, R. (2012). Health care costs associated with gestational diabetes mellitus among high-risk women: Results from a randomised trial. *BMC Pregnancy and Childbirth, 12*, 71. http://doi.org/10.1186/1471-2393-12-71

Krajmalnik-Brown, R., Ilhan, Z.-E., Kang, D.-W., & DiBaise, J. K. (2012). Effects of gut microbes on nutrient absorption and energy regulation. *Nutrition in Clinical Practice: Official Publication of the American Society for Parenteral and Enteral Nutrition, 27*(2), 201–214. http://doi.org/10.1177/0884533611436116

Kramer, C. K., Swaminathan, B., Hanley, A. J., Connelly, P. W., Sermer, M., Zinman, B., & Retnakaran, R. (2014). Vitamin D and parathyroid hormone status in pregnancy: Effect on insulin sensitivity, β-cell function, and gestational diabetes mellitus. *The Journal of Clinical Endocrinology & Metabolism, 99*(12), 4506–4513. https://doi.org/10.1210/jc.2014-2341

Krishnaveni, G. V., Hill, J. C., Veena, S. R., Bhat, D. S., Wills, A. K., Karat, C. L., … Fall,

C. H. (2009). Low plasma vitamin B12 in pregnancy is associated with gestational "diabesity" and later diabetes. *Diabetologia*, *52*(11), 2350–2358. https://doi.org/10.1007/s00125-009-1499-0

Lacroix, M., Battista, M.-C., Doyon, M., Houde, G., Ménard, J., Ardilouze, J.-L., … Perron, P. (2014). Lower vitamin D levels at first trimester are associated with higher risk of developing gestational diabetes mellitus. *Acta Diabetologica*, *51*(4), 609–616. https://doi.org/10.1007/s00592-014-0564-4

Landaas, S., & Pettersen, J. E. (1975). Clinical conditions associated with urinary excretion of 2-hydroxybutyric acid. *Scandinavian Journal of Clinical and Laboratory Investigation*, *35*(3), 259–266.

Langer, O., Yogev, Y., Most, O., & Xenakis, E. M. (2005). Gestational diabetes: The consequences of not treating. *American Journal of Obstetrics and Gynecology*, *192*(4), 989–997. https://doi.org/10.1016/j.ajog.2004.11.039

Lao, T. T., Ho, L.-F., Chan, B. C., & Leung, W.-C. (2006). Maternal age and prevalence of gestational diabetes mellitus. *Diabetes Care*, *29*(4), 948–949. https://doi.org/10.2337/DIACARE.29.04.06.DC05-2568

Larsen, N., Vogensen, F. K., van den Berg, F. W., Nielsen, D. S., Andreasen, A. S., Pedersen, B. K., … Munro, K. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS ONE*, *5*(2), e9085. http://doi.org/10.1371/journal .pone.0009085

Law, K. P., Mao, X., Han, T. L., & Zhang, H. (2017). 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. *Clinica Chimica Acta*, *465*, 53-71.

Law, K. P., Han, T. L., Mao, X., & Zhang, H. (2017a). 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. *Clinica Chimica Acta*, *468*, 126-139.

Lawson, E. J., & Rajaram, S. (1994). A transformed pregnancy: The psychosocial consequences of gestational diabetes. *Sociology of Health and Illness*, *16*(4), 536–562. http://doi.org/10.1111/1467-9566.ep11347644

Lecky, D. M., Hawking, M. K., & McNulty, C. A. (2014). Patients' perspectives on providing a stool sample to their GP: a qualitative study. Br J Gen Pract, 64(628), e684-e693.

Lee, P., Greenfield, J. R., Ho, K. K., & Fulham, M. J. (2010). A critical appraisal of the prevalence and metabolic significance of brown adipose tissue in adult humans. *American Journal of Physiology*, *299*(4), E601–E606.

Ley, S. H., Hanley, A. J., Retnakaran, R., Sermer, M., Zinman, B., & O'Connor, D. L. (2011). Effect of macronutrient intake during the second trimester on glucose metabolism later in pregnancy. *American Journal of Clinical Nutrition*, *94*(5), 1232–1240. https://doi.org/10.3945/ajcn.111.018861

Ley, S. H., Hanley, A. J., Sermer, M., Zinman, B., & O'Connor, D. L. (2013). Lower dietary vitamin E intake during the second trimester is associated with insulin resistance and hyperglycemia later in pregnancy. *European Journal of Clinical Nutrition*, *67*(11), 1154–1156. https://doi.org/10.1038/ejcn.2013.185

Li, L., Cheung, C., Ikram, M., Saw, S.-M., & Wong, T. (2013). Gestational diabetes mellitus (GDM) and changes in retinal microvasculature 6 months after delivery. *Investigative Ophthalmology & Visual Science*, *54*(15), 1561–1561.

Li, Q., Wu, X., Yang, S., Xie, M., & Shi, L. (2015). Association of dietary habits with gestational diabetes mellitus among Cantonese women. *Journal of Southern Medical University*, *35*(5), 772–776.

Li, Y., Ju, L., Hou, Z., Deng, H., Zhang, Z., Wang, L., … Zhang, Y. (2015). Screening, verification, and optimization of biomarkers for early prediction of cardiotoxicity based on metabolomics. *J Proteome Res*, *14*(6), 2437–2445. https://doi.org/10.1021/pr501116c

Liu, T., Li, J., Xu, F., Wang, M., Ding, S., Xu, H., & Dong, F. (2016). Comprehensive analysis of serum metabolites in gestational diabetes mellitus by UPLC/Q-TOF-MS. *Analytical and bioanalytical chemistry*, *408*(4), 1125-1135.

Lim, U., Ernst, T., Buchthal, S. D., Latch, M., Albright, C. L., Wilkens, L. R., … Le Marchand, L. (2011). Asian women have greater abdominal and visceral adiposity than Caucasian women with similar body mass index. *Nutrition and Diabetes*, *1*(5), e6. https://doi.org/10.1038/nutd.2011.2

Lloyd, A. J., Beckmann, M., Favé, G., Mathers, J. C., & Draper, J. (2011). Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. *British Journal of Nutrition*, *106*(6), 812–824. https://doi.org/10.1017/S0007114511001164

Loy, S. L., Chan, J. K., Wee, P. H., Colega, M. T., Cheung, Y. B., Godfrey, K. M., … Yap, F. (2017). Maternal circadian eating time and frequency are associated with blood glucose concentrations during pregnancy. *The Journal of Nutrition*, *147*(1), 70–77. https://doi.org/10.3945/jn.116.239392

Loy, S. L., Lek, N., Yap, F., Soh, S. E., Padmapriya, N., Tan, K. H., … Chan, J. K. (2015). Association of maternal vitamin D status with glucose tolerance and caesarean section in a multi-ethnic Asian cohort: The Growing Up in Singapore Towards Healthy Outcomes Study. *PLOS ONE*, *10*(11), e0142239. https://doi.org/10.1371/journal.pone.0142239

Luft, V. C., Schmidt, M. I., Pankow, J. S., Couper, D., Ballantyne, C. M., Hunter Young, J., & Duncan, B. B. (2013). Chronic inflammation role in the obesity-diabetes association: A case-cohort study. *Diabetology & Metabolic Syndrome*, *5*, 31.

Luoto, R., Laitinen, K., Nermes, M., & Isolauri, E. (2010). Impact of maternal probiotic-supplemented dietary counselling on pregnancy outcome and prenatal and postnatal growth: A double-blind, placebo-controlled study. *British Journal of Nutrition*, *103*(12), 1792–1799.

Magnusdottir, O. K., Landberg, R., Gunnarsdottir, I., Cloetens, L., Åkesson, B., Landin-Olsson, M., … Risérus, U. (2014). Plasma alkylresorcinols C17:0/C21:0 ratio, a biomarker of relative whole-grain rye intake, is associated to insulin sensitivity: A randomized study. *European Journal of Clinical Nutrition*, *68*(4), 453–458. http://doi.org/10.1038/ejcn.2014.12

Makgoba, M., Nelson, S. M., Savvidou, M., Messow, C.-M., Nicolaides, K., & Sattar, N. (2011). First-trimester circulating 25-hydroxyvitamin D levels and development of gestational diabetes mellitus. *Diabetes Care*, *34*(5), 1091–1093. https://doi.org/10.2337/dc10-2264

Makgoba, M., Savvidou, M., & Steer, P. (2012). 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, 119*(3), 276–282. https://doi.org/10.1111/j.1471-0528.2011.03156.x

Markovic, T. P., Muirhead, R., Overs, S., Ross, G. P., Louie, J. C., Kizirian, N., … Brand-Miller, J. C. (2016). Randomized controlled trial investigating the effects of a low–glycemic index diet on pregnancy outcomes in women at high risk of gestational diabetes mellitus: The GI Baby 3 Study. *Diabetes Care, 39*(1), 31–38. https://doi.org/10.2337/dc15-0572

Mastrogiannis, D., Spiliopoulos, M., Mulla, W., & Homko, C. (2009). Insulin resistance: The possible link between gestational diabetes mellitus and hypertensive disorders of pregnancy. *Current Diabetes Reports, 9*(4), 296–302.

McLeod, D. S., Warner, J. V., Henman, M., Cowley, D., Gibbons, K., McIntyre, H. D., & HAPO Study Cooperative Research Group. (2012). Associations of serum vitamin D concentrations with obstetric glucose metabolism in a subset of the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study cohort. *Diabetic Medicine, 29*(8), e199–e204. https://doi.org/10.1111/j.1464-5491.2011.03551.x

Meinila, J., Valkama, A., Koivusalo, S. B., Rönö, K., Kautiainen, H., Lindström, J., … Erkkola, M. (2017). Association between diet quality measured by the healthy food intake index and later risk of gestational diabetes: A secondary analysis of the RADIEL trial. *European Journal of Clinical Nutrition. 71*(4), 555-557. https://doi.org/10.1038/ejcn.2016.275

Menni, C., Zhai, G., MacGregor, A., Prehn, C., Römisch-Margl, W., Suhre, K., … Valdes, A. M. (2013). Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics, 9*(2), 506–514. https://doi.org/10.1007/s11306-012-0469-6

Metzger, B. E., Buchanan, T. A., Coustan, D. R., de Leiva, A., Dunger, D. B., Hadden, D. R., … Zoupas, C. (2007). Summary and recommendations of the Fifth International Workshop-Conference on Gestational Diabetes Mellitus. *Diabetes Care, 30*(Supplement 2), S251–S260.

Meyer, B., Calvert, D., & Moses, R. (1996). Free fatty acids and gestational diabetes mellitus. *The Australian & New Zealand Journal of Obstetrics & Gynaecology, 36*(3), 255–257.

Michelucci, A., Cordes, T., Ghelfi, J., Pailot, A., Reiling, N., Goldmann, O., … Hiller, K. (2013). Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. *Proceedings of the National Academy of Sciences, 110*(19), 7820–7825. http://doi.org/10.1073/pnas.1218599110

Ministry of Health. (2008). *Food and nutrition guidelines for healthy pregnant and breastfeeding women: A background paper*. Wellington, New Zealand: Author.

Ministry of Health. (2014). *Eating for healthy pregnant women*. Retrieved from https://www.healthed.govt.nz/system/files/resource-files/HE1805-Eating for healthy pregnant women.pdf

Miyazaki, Y., Glass, L., Triplitt, C., Wajcberg, E., Mandarino, L. J., & DeFronzo, R. A. (2002). Abdominal fat distribution and peripheral and hepatic insulin resistance in type 2 diabetes mellitus. *American Journal of Physiology, 283*(6), E1135–E1143.

Mohan, V., Radhika, G., Sathya, R. M., Tamil, S. R., Ganesan, A., & Sudha, V. (2009). Dietary carbohydrates, glycaemic load, food groups and newly detected type 2 diabetes

among urban Asian Indian population in Chennai, India (Chennai Urban Rural Epidemiology Study 59). *The British Journal of Nutrition*, *102*(10), 1498–1506. http://doi.org/10.1017/S0007114509990468

Mojibian, M., Soheilykhah, S., Fallah Zadeh, M. A., & Jannati Moghadam, M. (2015). The effects of vitamin D supplementation on maternal and neonatal outcome: A randomized clinical trial. *Iranian Journal of Reproductive Medicine*, *13*(11), 687–696.

Mora-Cubillos, X., Tulipani, S., Garcia-Aloy, M., Bulló, M., Tinahones, F. J., & Andres-Lacueva, C. (2015). Plasma metabolomic biomarkers of mixed nuts exposure inversely correlate with severity of metabolic syndrome. *Molecular Nutrition & Food Research*, *59*(12), 2480–2490. http://doi.org/10.1002/mnfr.201500549

MRC Vitamin Study Research Group. (1991). Prevention of neural tube defects: Results of the Medical Research Council Vitamin Study. *The Lancet*, *338*(8760), 131–137. https://doi.org/10.1016/0140-6736(91)90133-A

Mujezinovic, F., & Alfirevic, Z. (2007). Procedure-related complications of amniocentesis and chorionic villous sampling. *Obstetrics & Gynecology*, *110*(3), 687–694. http://doi.org/10.1097/01.AOG.0000278820.54029.e3

Murphy, A., Shamshirsaz, A., Markovic, D., Ostlund, R., & Koos, B. (2016). Urinary excretion of *myo*-inositol and d-*chiro*-inositol in early pregnancy is enhanced in gravidas with gestational diabetes mellitus. *Reproductive Sciences*, *23*(3), 365–371. https://doi.org/10.1177/1933719115602767

Nair, R., & Maseeh, A. (2012). Vitamin D: The" sunshine" vitamin. *Journal of Pharmacology and Pharmacotherapeutics*, *3*(2), 118.

Nanri, A., Mizoue, T., Noda, M., Takahashi, Y., Kato, M., Inoue, M., … Japan Public Health Center-Based Prospective Study Group. (2010). Rice intake and type 2 diabetes in Japanese men and women: The Japan Public Health Center-Based Prospective Study. *The American Journal of Clinical Nutrition*, *92*(6), 1468–1477. http://doi.org/10.3945/ajcn.2010.29512

National Diabetes Data Group. (1979). Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes*, *28*(12), 1039–1057.

Neeland, I. J., Turer, A. T., Ayers, C. R., Powell-Wiley, T. M., Vega, G. L., Farzaneh-Far, R., … de Lemos, J. A. (2012). Dysfunctional adiposity and the risk of prediabetes and type 2 diabetes in obese adults. *JAMA*, *308*(11), 1150–1159.

Newby, P. K., Muller, D., Hallfrisch, J., Andres, R., & Tucker, K. L. (2004). Food patterns measured by factor analysis and anthropometric changes in adults. *The American Journal of Clinical Nutrition*, *80*(2), 504–513.

Nicholson, J. K., & Lindon, J. C. (2008). Systems biology: metabonomics. *Nature, 455* (7216), 1054-1056.

Nobles, C. J., Markenson, G., & Chasan-Taber, L. (2015). Early pregnancy vitamin D status and risk for adverse maternal and infant outcomes in a bi-ethnic cohort: The Behaviors Affecting Baby and You (B.A.B.Y.) study. *The British Journal of Nutrition*, *114*(12), 2116–2128. https://doi.org/10.1017/S0007114515003980

O'Gorman, A., Gibbons, H., & Brennan, L. (2013). Metabolomics in the identification of biomarkers of dietary intake. *Computational and Structural Biotechnology Journal*, *4*(5), e201301004. https://doi.org/10.5936/csbj.201301004

O'Sullivan, A., Gibney, M. J., & Brennan, L. (2011). Dietary intake patterns are reflected in metabolomic profiles: Potential role in dietary assessment studies. *American Journal of Clinical Nutrition, 93*(2), 314–321. https://doi.org/10.3945/ajcn.110.000950

Oliver, S. G., Winson, M. K., Kell, D. B., & Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends Biotechnol, 16*(9), 373-378

Omidvar, S., Sharbatdaran, M., Tilaki, K., Pouramir, M., Amiri, F., & Basirat, Z. (2013). Comparison of the serum iron, ferritin levels and total iron-binding capacity between pregnant women with and without gestational diabetes. *Journal of Natural Science, Biology and Medicine, 4*(2), 302. https://doi.org/10.4103/0976-9668.116977

Osler, M., Heitmann, B. L., Gerdes, L. U., Jùrgensen, L. M., & Schroll, M. (2001). Dietary patterns and mortality in Danish men and women: A prospective observational study. *British Journal of Nutrition, 85*(2), 219–225. http://doi.org/10.1079/BJN2000240

Osorio-Yáñez, C., Qiu, C., Gelaye, B., Enquobahrie, D. A., & Williams, M. A. (2016). Risk of gestational diabetes mellitus in relation to maternal dietary calcium intake. *Public Health Nutrition, 20*(6), 1082-1089. https://doi.org/10.1017/S1368980016002974

Pamungkas, A. D., Park, C., Lee, S., Jee, S. H., & Park, Y. H. (2016). High resolution metabolomics to discriminate compounds in serum of male lung cancer patients in South Korea. *Respir Res 17*(1), 8. https://doi.org/10.1186/s12931-016-0419-3

Pan, Z., & Raftery, D. (2007). Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Analytical and Bioanalytical Chemistry, 387*(2), 525–527.

Papadimitriou, A., King, A. J., Jones, P. M., & Persaud, S. J. (2007). Anti-apoptotic effects of arachidonic acid and prostaglandin E2 in pancreatic beta-cells. *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology, 20*(5), 607–616. http://doi.org/10.1159/000107544

Parada, J., & Aguilera, J. M. (2007). Food microstructure affects the bioavailability of several nutrients. *Journal of Food Science, 72*(2), R21–R32.

Park, H.-J., Lee, J., Kim, J.-M., Lee, H. A., Kim, S.-H., & Kim, Y. (2013). A study of snack consumption, night-eating habits, and nutrient intake in gestational diabetes mellitus. *Clinical Nutrition Research, 2*(1), 42–51. https://doi.org/10.7762/cnr.2013.2.1.42

Park, S., Kim, M.-Y., Baik, S. H., Woo, J.-T., Kwon, Y. J., Daily, J. W., … Kim, S.-H. (2013). Gestational diabetes is associated with high energy and saturated fat intakes and with low plasma visfatin and adiponectin levels independent of prepregnancy BMI. *European Journal of Clinical Nutrition, 67*(2), 196–201. https://doi.org/10.1038/ejcn.2012.207

Park, S., Park, J. Y., Lee, J. H., & Kim, S.-H. (2015). Plasma levels of lysine, tyrosine, and valine during pregnancy are independent risk factors of insulin resistance and gestational diabetes. *Metabolic Syndrome and Related Disorders, 13*(2), 64–70. http://doi.org/10.1089/met.2014.0113

Park, S., Yoon, H.-K., Ryu, H.-M., Han, Y. J., Lee, S. W., Park, B. K., … Kim, S.-H. (2014). Maternal vitamin D deficiency in early pregnancy is not associated with gestational diabetes mellitus development or pregnancy outcomes in Korean pregnant women in a prospective study. *Journal of Nutritional Science and Vitaminology, 60*(4), 269–275.

Park, Y.-W., Allison, D. B., Heymsfield, S. B., & Gallagher, D. (2001). Larger amounts of

visceral adipose tissue in Asian Americans. *Obesity Research*, *9*(7), 381–387. https://doi.org/10.1038/oby.2001.49

Parlea, L., Bromberg, I. L., Feig, D. S., Vieth, R., Merman, E., & Lipscombe, L. L. (2012). Association between serum 25-hydroxyvitamin D in early pregnancy and risk of gestational diabetes mellitus. *Diabetic Medicine*, *29*(7), e25–e32. https://doi.org/10.1111/j.1464-5491.2011.03550.x

Patel, P. S., Cooper, A. J., O'Connell, T. C., Kuhnle, G. G., Kneale, C. K., Mulligan, A. M., … Forouhi, N. G. (2014). Serum carbon and nitrogen stable isotopes as potential biomarkers of dietary intake and their relation with incident type 2 diabetes: The EPIC-Norfolk study. *American Journal of Clinical Nutrition*, *100*(2), 708–718. http://doi.org/10.3945/ajcn.113.068577

Patti, G. J., Yanes, O., & Siuzdak, G. (2012). Metabolomics: the apogee of the omic triology. Nature reviews. *Molecular cell biology*, *13*(4), 263.

Pedersen, V., Gudmundsdottir, H., Nielsen, H., Hyotylainen, T., Nielsen, T., Jensen, B., … Pedersen, O. (2016). Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature*, *535*, 376–381.

Pinto, J., Almeida, L. M., Martins, A. S., Duarte, D., Barros, A. S., Galhano, E., … Gil, A. M. (2015). Prediction of gestational diabetes through nmr metabolomics of maternal blood. *Journal of Proteome Research*, *14*(6), 2696–2706. https://doi.org/10.1021/acs.jproteome.5b00260

Pleskačová, A., Bartáková, V., Pácal, L., Kuricová, K., Bělobrádková, J., Tomandl, J., & Kaňková, K. (2015). Vitamin D status in women with gestational diabetes mellitus during pregnancy and postpartum. *BioMed Research International*, *2015*, 1–7. https://doi.org/10.1155/2015/260624

Poston, L., Briley, A. L., Barr, S., Bell, R., Croker, H., Coxon, K., ... Khazaezadeh, N. (2013). Developing a complex intervention for diet and activity behaviour change in obese pregnant women (the UPBEAT trial); assessment of behavioural change and process evaluation in a pilot randomised controlled trial. *BMC pregnancy and childbirth*, *13*(1), 148.

Puhl, R. M., & Latner, J. D. (2007). Stigma, obesity, and the health of the nation's children. *Psychological Bulletin*, *133*(4), 557-580.

Qiu, C., Frederick, I. ., Zhang, C., Sorensen, T. ., Enquobahrie, D. A., & Williams, M. A. (2011). Risk of gestational diabetes mellitus in relation to maternal egg and cholesterol intake. *American Journal of Epidemiology*, *173*(6), 649–658.

Radesky, J. S., Oken, E., Rifas-Shiman, S. L., Kleinman, K. P., Rich-Edwards, J. W., & Gillman, M. W. (2007). Diet during early pregnancy and development of gestational diabetes. *Paediatric and Perinatal Epidemiology*, *22*(1), 47–59. https://doi.org/10.1111/j.1365-3016.2007.00899.x

Rana, J. S., Li, T. Y., Manson, J. E., & Hu, F. B. (2007). Adiposity compared with physical inactivity and risk of type 2 diabetes in women. *Diabetes Care*, *30*(1), 53–58.

Rasmussen, L. G., Savorani, F., Larsen, T. M., Dragsted, L. O., Astrup, A., & Engelsen, S. B. (2011). Standardization of factors that influence human urine metabolomics. *Metabolomics*, *7*(1), 71–83. http://doi.org/10.1007/s11306-010-0234-7

Raterink, R. J., Lindenburg, P. W., Vreeken, R. J., Ramautar, R., & Hankemeier, T. (2014).

Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics. *TrAC Trends in Analytical Chemistry, 61*, 157-167.

Rathnayake, K., Madushani, P., & Silva, K. (2012). Use of dietary diversity score as a proxy indicator of nutrient adequacy of rural elderly people in Sri Lanka. *BMC Research Notes*, *5*, 469. http://doi.org/10.1186/1756-0500-5-469

Rawal, S., Hinkle, S. N., Bao, W., Zhu, Y., Grewal, J., Albert, P. S., … Zhang, C. (2017). A longitudinal study of iron status during pregnancy and the risk of gestational diabetes: Findings from a prospective, multiracial cohort. *Diabetologia*, *60*(2), 249–257. https://doi.org/10.1007/s00125-016-4149-3

Reedy, J., Wirfält, E., Flood, A., Mitrou, P. N., Krebs-Smith, S. M., Kipnis, V., ... & Subar, A. F. (2009). Comparing 3 dietary pattern methods—cluster analysis, factor analysis, and index analysis—with colorectal cancer risk: the NIH–AARP Diet and Health Study. *American Journal of Epidemiology*, *171*(4), 479-487.

Ritov, V., Menshikova, E., He, J., Ferrell, R., Goodpaster, B., & Kelley, D. (2005). Deficiency of subsarcolemmal mitochondria in obesity and type 2 diabetes. *Diabetes*, *54*(1), 8–14.

Robert, M. F., Neff, R. K., Hubbell, J. P., Taeusch, H. W., & Avery, M. E. (1976). Association between maternal diabetes and the respiratory-distress syndrome in the newborn. *The New England Journal of Medicine*, *294*(7), 357–360. https://doi.org/10.1056/NEJM197602122940702

Rodriguez, A., García-Esteban, R., Basterretxea, M., Lertxundi, A., Rodríguez-Bernal, C., Iñiguez, C., … Morales, E. (2015). Associations of maternal circulating 25-hydroxyvitamin D3 concentration with pregnancy and birth outcomes. *BJOG: An International Journal of Obstetrics & Gynaecology*, *122*(12), 1695–1704. https://doi.org/10.1111/1471-0528.13074

Rothwell, J. A., Fillâtre, Y., Martin, J.-F., Lyan, B., Pujos-Guillot, E., Fezeu, L., … Kristal, B. (2014). New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. *PLoS ONE*, *9*(4), e93474. https://doi.org/10.1371/journal.pone.0093474

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473–489.

Rush, E. C., Goedecke, J. H., Jennings, C., Micklesfield, L., Dugas, L., Lambert, E. V, & Plank, L. D. (2007). BMI, fat and muscle differences in urban women of five ethnicities from two countries. *International Journal of Obesity*, *31*(8), 1232–1239. https://doi.org/10.1038/sj.ijo.0803576

Sachse, D., Sletner, L., Mørkrid, K., Jenum, A. K., Birkeland, K. I., Rise, F., … Berg, J. P. (2012). Metabolic changes in urine during and after pregnancy in a large, multiethnic population-based cohort study of gestational diabetes. *PloS One*, *7*(12), e52399. https://doi.org/10.1371/journal.pone.0052399

Sahariah, S. A., Potdar, R. D., Gandhi, M., Kehoe, S. H., Brown, N., Sane, H., … Fall, C. H. (2016). A daily snack containing leafy green vegetables, fruit, and milk before and during pregnancy prevents gestational diabetes in a randomized, controlled trial in Mumbai, India. *Journal of Nutrition*, *146*(7), 1453S–1460S. https://doi.org/10.3945/jn.115.223461

Salmenhaara, M., Uusitalo, L. U., Kronberg-Kippilä, C., Sinkko, H., Ahonen, S., …

Virtanen, S. . (2010). Diet and weight gain characteristics of pregnant women with gestational diabetes. *European Journal of Clinical Nutrition*, *64*, 1433–1440.

Salmerón, J., Manson, J. E., Stampfer, M. J., Colditz, G. A., Wing, A. L., & Willett, W. C. (1997). Dietary fiber, glycemic load, and risk of non-insulin-dependent diabetes mellitus in women. *JAMA*, *277*(6), 472–477.

Santamaria, A., Corrado, F., Baviera, G., Carlomagno, G., Unfer, V., & D'Anna, R. (2016). Second trimester amniotic fluid myo-inositol concentrations in women later developing gestational diabetes mellitus or pregnancy-induced hypertension. *The Journal of Maternal-Fetal & Neonatal Medicine*, *29*(14), 2245–2247. https://doi.org/10.3109/14767058.2015.1081886

Saude, E. J., Adamko, D., Rowe, B. H., Marrie, T., & Sykes, B. D. (2007). Variation of metabolites in normal human urine. *Metabolomics*, *3*(4), 439–451. http://doi.org/10.1007/s11306-007-0091-1

Savvidou, M., Akolekar, R., Samaha, R., Masconi, A., & Nicolaides, K. (2011). Maternal serum 25-hydroxyvitamin D levels at 11+0-13+6 weeks in pregnant women with diabetes mellitus and in those with macrosomic neonates. *BJOG: An International Journal of Obstetrics & Gynaecology*, *118*(8), 951–955. https://doi.org/10.1111/j.1471-0528.2011.02982.x

Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L. O., Draper, J., … Wishart, D. S. (2014). The food metabolome: A window over dietary exposure. *American Journal of Clinical Nutrition*, *99*(6), 1286–1308. https://doi.org/10.3945/ajcn.113.076133

Schaefer, A., Jensen, E., Bossart, G., & Reif, J. (2014). Hair mercury concentrations and fish consumption patterns in Florida residents. *International Journal of Environmental Research and Public Health*, *11*(7), 6709–6726. http://doi.org/10.3390/ijerph110706709

Schmidt, J. A., Rinaldi, S., Ferrari, P., Carayol, M., Achaintre, D., Scalbert, A., … Travis, R. C. (2015). Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. *American Journal of Clinical Nutrition*, *102*(6), 1518–1526. https://doi.org/10.3945/ajcn.115.111989

Schneuer, F. J., Roberts, C. L., Guilbert, C., Simpson, J. M., Algert, C. S., Khambalia, A. Z., … Nassar, N. (2014). Effects of maternal serum 25-hydroxyvitamin D concentrations in the first trimester on subsequent pregnancy outcomes in an Australian population. *American Journal of Clinical Nutrition*, *99*(2), 287–295. https://doi.org/10.3945/ajcn.113.065672

Schoenaker, D. A., Soedamah-Muthu, S. S., Callaway, L. K., & Mishra, G. D. (2015). Pre-pregnancy dietary patterns and risk of gestational diabetes mellitus: Results from an Australian population-based prospective cohort study. *Diabetologia*, *58*(12), 2726–2735. https://doi.org/10.1007/s00125-015-3742-1

Sebire, N. J., Jolly, M., Harris, J. P., Wadsworth, J., Joffe, M., Beard, R. W., … Robinson, S. (2001). Maternal obesity and pregnancy outcome: A study of 287 213 pregnancies in London. *International Journal of Obesity*, *25*(8), 1175–1182. https://doi.org/10.1038/sj.ijo.0801670

Seidell, J., Björntorp, P., Sjöström, L., Kvist, H., & Sannerstedt, R. (1990). Visceral fat accumulation in men is positively associated with insulin, glucose, and c-peptide levels, but negatively with testosterone levels. *Metabolism*, *39*(9), 897–901.

Shahgheibi, S., Farhadifar, F., & Pouya, B. (2016). The effect of vitamin D supplementation on gestational diabetes in high-risk women: Results from a randomized placebo-controlled trial. *Journal of Research in Medical Sciences*, *21*(1), 2. https://doi.org/10.4103/1735-1995.175148

Shahwar, M. K., El-Ghorab, A. H., Anjum, F. M., Butt, M. S., Hussain, S., & Nadeem, M. (2012). Characterization of coriander (*Coriandrum sativum* L.) seeds and leaves: Volatile and non volatile extracts. *International Journal of Food Properties*, *15*(4), 736–747. http://doi.org/10.1080/10942912.2010.500068

Shai, I., Jiang, R., Manson, J. E., Stampfer, M. J., Willett, W. C., Colditz, G. A., & Hu, F. B. (2006). Ethnicity, obesity, and risk of type 2 diabetes in women: A 20-year follow-up study. *Diabetes Care*, *29*(7), 1585–1590. https://doi.org/10.2337/dc06-0057

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591-611.

Shen, J., Kai, J., Tang, Y., Zhang, L., Su, S., & Duan, J.-A. (2016). The chemical and biological properties of Euphorbia kansui. *The American Journal of Chinese Medicine, 44*(02), 253-273. http://doi.org/10.1142/S0192415X16500154

Shen, V., Siderius, D., Krekelberg, W., & Hatch, H. (2017). *NIST standard reference database 1A v14 | NIST*. Retrieved from https://www.nist.gov/srd/nist-standard-reference-database-1a-v14

Shin, D., Lee, K., & Song, W. (2015). Dietary patterns during pregnancy are associated with risk of gestational diabetes mellitus. *Nutrients*, *7*(11), 9369–9382. https://doi.org/10.3390/nu7115472

Silverman, B. L., Metzger, B. E., Cho, N. H., & Loeb, C. A. (1995). Impaired glucose tolerance in adolescent offspring of diabetic mothers. Relationship to fetal hyperinsulinism. *Diabetes Care*, *18*(5), 611–7. https://doi.org/10.2337 /DIACARE.18.5.611

Silverman, B. L., Rizzo, T. A., Cho, N. H., & Metzger, B. E. (1998). Long-term effects of the intrauterine environment. The Northwestern University Diabetes in Pregnancy Center. *Diabetes Care*, *21 Suppl 2*, B142-149.

Silverman, M. E., Reichenberg, A., Savitz, D. A., Cnattingius, S., Lichtenstein, P., Hultman, C. M., … Sandin, S. (2017). The risk factors for postpartum depression: A population-based study. *Depression and Anxiety*, *34*(2), 178–187. https://doi.org/10.1002/da.22597

Smart, K. F., Aggio, R. B., Van Houtte, J. R., & Villas-Bôas, S. G. (2010). Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography-mass spectrometry. *Nature Protocols*, *5*(10), 1709–1729. http://doi.org/10.1038/nprot.2010.108

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, *78*(3), 779–787. http://doi.org/10.1021/ac051437y

Sparks, L., Xie, H., Koza, R., Mynatt, R., Hulver, M., Bray, G., & Smith, S. (2005). A high-fat diet coordinately downregulates genes required for mitochondrial oxidative phosphorylation in skeletal muscle. *Diabetes*, *54*(7), 1926–1933.

Stein, S. E. (1999). An integrated method for spectrum extraction and compound

identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry*, *10*(8), 770–781. http://doi.org/10.1016/S1044-0305(99)00047-1

Stradling, C., Thomas, G. N., Hemming, K., Frost, G., Garcia-Perez, I., Redwood, S., & Taheri, S. (2016). Randomised controlled pilot study to assess the feasibility of a Mediterranean Portfolio dietary intervention for cardiovascular risk reduction in HIV dyslipidaemia: A study protocol. *BMJ Open*, *6*(2), e010821. https://doi.org/10.1136/bmjopen-2015-010821

Sugiyama, M., Tang, A. C., Wakaki, Y., & Koyama, W. (2003). Glycemic index of single and mixed meal foods among common Japanese foods with white rice as a reference food. *European Journal of Clinical Nutrition*, *57*(6), 743–752. http://doi.org/10.1038/sj.ejcn.1601606

Suhre, K., Meisinger, C., Döring, A., Altmaier, E., Belcredi, P., Gieger, C., … Delzenne, N. (2010). Metabolic footprint of diabetes: A multiplatform metabolomics study in an epidemiological setting. *PLoS ONE*, *5*(11), e13953. http://doi.org/10.1371/journal.pone.0013953

Sukumar, N., Venkataraman, H., Wilson, S., Goljan, I., Selvamoni, S., Patel, V., & Saravanan, P. (2016). Vitamin B12 status among pregnant women in the UK and its association with obesity and gestational diabetes. *Nutrients*, *8*(12), E768. https://doi.org/10.3390/nu8120768

Sulek, K., Han, T.-L., Villas-Boas, S. G., Wishart, D. S., Soh, S.-E., Kwek, K., … Baker, P. N. (2014). Hair metabolomics: Identification of fetal compromise provides proof of concept for biomarker discovery. *Theranostics*, *4*(9), 953–959. http://doi.org/10.7150/thno.9265

Thompson, J. M., Wall, C., Becroft, D. M., Robinson, E., Wild, C. J., & Mitchell, E. A. (2010). Maternal dietary patterns in pregnancy and the association with small-for-gestational-age infants. *British Journal of Nutrition*, *103*(11), 1665–1673. http://doi.org/10.1017/S0007114509993606

Thurnham, D. I., McCabe, L. D., Haldar, S., Wieringa, F. T., Northrop-Clewes, C. A., & McCabe, G. P. (2010). Adjusting plasma ferritin concentrations to remove the effects of subclinical inflammation in the assessment of iron deficiency: A meta-analysis. *The American Journal of Clinical Nutrition*, 92(3), 546-555. doi:10.3945/ajcn.2010.29284

Tryggvadottir, E. A., Medek, H., Birgisdottir, B. E., Geirsson, R. T., & Gunnarsdottir, I. (2016). Association between healthy maternal dietary pattern and risk for gestational diabetes mellitus. *European Journal of Clinical Nutrition*, *70*(2), 237–242. https://doi.org/10.1038/ejcn.2015.145

United Nations Office on Drugs and Crimes. (2014). *Guidelines for testing drugs under international control in hair, sweat and oral fluid*. New York, NY: Author.

van Wijk, R., & van Solinge, W. W. (2005). The energy-less red blood cell is lost: erythrocyte enzyme abnormalities of glycolysis. *Blood*, *106*(13), 4034–4042. http://doi.org/10.1182/blood-2005-04-1622

Vázquez-Fresno, R., Llorach, R., Urpi-Sarda, M., Lupianez-Barbero, A., Estruch, R., Corella, D., … Andres-Lacueva, C. (2015). Metabolomic pattern analysis after Mediterranean diet intervention in a nondiabetic population: A 1- and 3-year follow-up in the PREDIMED study. *Journal of Proteome Research*, *14*(1), 531–540.

https://doi.org/10.1021/pr5007894

Venn-Watson, S. K., Parry, C., Baird, M., Stevenson, S., Carlin, K., Daniels, R., … Jensen, E. D. (2015). Increased dietary intake of saturated fatty acid heptadecanoic acid (C17:0) associated with decreasing ferritin and alleviated metabolic syndrome in dolphins. *PLOS ONE*, *10*(7), e0132117. http://doi.org/10.1371/journal.pone.0132117

Villas-Bôas, S. G. (2007). Sampling and sample preparation. In S. G. Villas-Bôas, U. Roessner, M. A. Hansen, J. Smedsgaard, & J. Nielsen (Eds.), *Metabolome analysis: An introduction* (pp. 39–82). Hoboken, NJ: Wiley-Interscience.

Villas-Bôas, S. G., Smart, K. F., Sivakumaran, S., & Lane, G. A. (2011). Alkylation or silylation for analysis of amino and non-amino organic acids by GC-MS? *Metabolites*, *1*(1), 3–20. http://doi.org/10.3390/metabo1010003

Villegas, R., Liu, S., Gao, Y.-T., Yang, G., Li, H., Zheng, W., & Shu, X. O. (2007). Prospective study of dietary carbohydrates, glycemic index, glycemic load, and incidence of type 2 diabetes mellitus in middle-aged Chinese women. *Archives of Internal Medicine*, *167*(21), 2310–2316. http://doi.org/10.1001/archinte.167.21.2310

Wanders, R. J., Komen, J., & Kemp, S. (2011). Fatty acid omega-oxidation as a rescue pathway for fatty acid oxidation disorders in humans. *FEBS Journal*, *278*(2), 182–194. http://doi.org/10.1111/j.1742-4658.2010.07947.x

Wang, O., Nie, M., Hu, Y. Y., Zhang, K., Li, W., Ping, F., … Xing, X. P. (2012). Association between vitamin D insufficiency and the risk for gestational diabetes mellitus in pregnant Chinese women. *Biomedical and Environmental Sciences: BES*, *25*(4), 399–406. https://doi.org/10.3967/0895-3988.2012.04.004

Wang, T. J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P., McCabe, E., … Gerszten, R. E. (2011). Metabolite profiles and the risk of developing diabetes. *Nature Medicine*, *17*(4), 448–453. http://doi.org/10.1038/nm.2307

Wang, Y., Rimm, E. B., Stampfer, M. J., Willett, W. C., & Hu, F. B. (2005). Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. *American Society for Clinical Nutrition*, *81*(3), 555–563.

Want, E., O'Maille, G., Smith, C., Brandon, T., Uritboonthai, W., Qin, C., … Siuzdak, G. (2005). Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Analytical Chemistry*, *78*(3), 743-752. http://doi.org/10.1021/AC051312T

Whitelaw, D. C., Scally, A. J., Tuffnell, D. J., Davies, T. J., Fraser, W. D., Bhopal, R. S., … Lawlor, D. A. (2014). Associations of circulating calcium and 25-hydroxyvitamin d with glucose metabolism in pregnancy: A cross-sectional study in European and South Asian women. *The Journal of Clinical Endocrinology & Metabolism*, *99*(3), 938–946. https://doi.org/10.1210/jc.2013-2896

Williamson, J. R., Wałajtys-Rode, E., & Coll, K. E. (1979). Effects of branched chain alpha-ketoacids on the metabolism of isolated rat liver cells. I. Regulation of branched chain alpha-ketoacid metabolism. *The Journal of Biological Chemistry*, *254*(22), 11511–11520.

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., … Scalbert, A. (2013). HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Research*, *41*(Database issue), D801-807. https://doi.org/10.1093/nar/gks1065

Wolf, M., Sandler, L., Hsu, K., Vossen-Smirnakis, K., Ecker, J. L., & Thadhani, R. (2003). First-trimester c-reactive protein and subsequent gestational diabetes. *Diabetes Care*, *26*(3), 819-824.

Wolk, A., Vessby, B., Ljung, H., & Barrefors, P. (1998). Evaluation of a biological marker of dairy fat intake. *The American Journal of Clinical Nutrition*, *68*(2), 291–295.

Xia, J., Psychogios, N., Young, N., & Wishart, D. S. (2009). MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Research*, *37*(suppl 2), W652-W660. http://doi.org/10.1093/nar/gkp356

Xu, J., Yang, S., Cai, S., Dong, J., Li, X., & Chen, Z. (2010). Identification of biochemical changes in lactovegetarian urine using 1H NMR spectroscopy and pattern recognition. *Analytical and Bioanalytical Chemistry*, *396*(4), 1451–1463. https://doi.org/10.1007/s00216-009-3338-z

Yap, C., Cheung, N. W., Gunton, J. E., Athayde, N., Munns, C. F., Duke, A., & McLean, M. (2014). Vitamin D supplementation and the effects on glucose metabolism during pregnancy: A randomized controlled trial. *Diabetes Care*, *37*(7), 1837–1844. https://doi.org/10.2337/dc14-0155

Yin, P., Lehmann, R., & Xu, G. (2015). Effects of pre-analytical processes on blood samples used in metabolomics studies. *Analytical and Bioanalytical Chemistry*, *407*(17), 4879–4892. http://doi.org/10.1007/s00216-015-8565-x

Yin, P., Peter, A., Franken, H., Zhao, X., Neukamm, S., Rosenbaum, L., … Lehmann, R. (2013). Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. *Clinical Chemistry*, *59*(5), 833–845. http://doi.org/10.1373/clinchem .2012.199257

Zein, S., Rachidi, S., Awada, S., Osman, M., Al-Hajje, A., Shami, N., … Hininger-Favier, I. (2015). High iron level in early pregnancy increased glucose intolerance. *Journal of Trace Elements in Medicine and Biology*, *30*, 220–225. https://doi.org/10.1016/j.jtemb.2014.09.004

Zhang, C., Qiu, C., Hu, F. B., David, R. M., van Dam, R. M., Bralley, A., & Williams, M. A. (2008). Maternal plasma 25-hydroxyvitamin D concentrations and the risk for gestational diabetes mellitus. *PloS One*, *3*(11), e3753. https://doi.org/10.1371/journal.pone.0003753

Zhang, C., Schulze, M. B., Solomon, C. G., & Hu, F. B. (2006). A prospective study of dietary patterns, meat intake and the risk of gestational diabetes mellitus. *Diabetologia*, *49*(11), 2604–2613. https://doi.org/10.1007/s00125-006-0422-1

Zhang, L., Shu, X., Ding, A., Yu, L., Tang, Y., Duan, J., … Shen, X. (2009). LC–DAD–ESI-MS–MS separation and chemical characterization of the inflammatory fraction of the roots of Euphorbia kansui. *Chromatographia*, *70*(5-6), 805–810. http://doi.org/10.1365/s10337-009-1258-x

Zheng, Y., Li, Y., Qi, Q., Hruby, A., Manson, J. E., Willett, W. C., … Qi, L. (2016). Cumulative consumption of branched-chain amino acids and incidence of type 2 diabetes. *International Journal of Epidemiology*, *45*(5), 1482-1492. dyw143. http://doi.org/10.1093/ije/dyw143

Zhou, J., Su, L., Liu, M., Liu, Y., Cao, X., Wang, Z., & Xiao, H. (2014). Associations between 25-hydroxyvitamin D levels and pregnancy outcomes: A prospective observational study in southern China. *European Journal of Clinical Nutrition*, *68*(8),

925–930. https://doi.org/10.1038/ejcn.2014.99

Zhu, B., Ge, X., Huang, K., Mao, L., Yan, S., Xu, Y., … Tao, F. (2016). Folic acid supplement intake in early pregnancy increases risk of gestational diabetes mellitus: Evidence from a prospective cohort study. *Diabetes Care*, *39*(3), e36-37. https://doi.org/10.2337/dc15-2389

Zuhur, S., Erol, R., Kuzu, I., & Altuntas, Y. (2013). The relationship between low maternal serum 25-hydroxyvitamin D levels and gestational diabetes mellitus according to the severity of 25-hydroxyvitamin D deficiency. *Clinics*, *68*(5), 658–664. https://doi.org/10.6061/clinics/2013(05)13

Zyznar, E. S. (1981). A rationale for the application of trasylol as a protease inhibitor in radioimmunoassay. *Life Sciences*, *28*(17), 1861–1866. http://doi.org/10.1016/0024-3205(81)90291-5