METABOLIC PHENOTYPING AND METAGENOMIC ANALYSIS OF DEVELOPING INFANTS

Thesis submitted by

Frances Jackson

For the degree of Doctor of Philosophy of Imperial College London

Supervisors:
Professor Jeremy Nicholson
Professor Neena Modi
Professor Julian Marchesi

Division of Computational and Systems Medicine

Department of Surgery and Cancer

Faculty of Medicine

Imperial College London

2016

ABSTRACT

Early life experiences, including mode of delivery and nutrition during the neonatal period, have been proven to have an impact on health in later life. Studying human metabolic development has major implications for understanding the aetiology and risk of disease, including metabolic syndrome.

Initially, a sample preparation protocol was developed and optimised using metabonomic procedures for studying urine and faeces from infants, to accommodate for limited sample volume and to take into account the compositional differences between adult and infant biofluids. This primarily indicated that age is an important variable that contributes to the metabolic profile of biofluids.

Faecal metabonomics is fast becoming a useful tool for defining interactions among host, microbial communities and nutritional interventions. Infant development trajectory was assessed through analysis of faecal metabolic profiling by ^{1}H NMR. A large non-clinical cohort longitudinal study was obtained; 1802 faecal samples from 524 infants at 6 time points from 4 days to 730 days postpartum. Furthermore, ^{1}H NMR, UPLC-MS and metagenomic phenotyping techniques was performed on urine (n=278) and faecal (n=308) samples from 150 infants born term or preterm (<37 wks gestational age). This multi-omics approach provided further demonstration of contribution of microbial co-metabolites to infant metabolism early in life and therefore the potential impact on overall health.

This PhD project was able to identify certain metabolic pathways which were shown to be different in relation to gestation age as well as postnatal age, mode of delivery, BMI status and nutrition. In particular, choline and methylamine derivatives (e.g. betaine, trimethylamine), short chain fatty acids (SCFA) and amino acids related to nutrition and the gut microbiome functionality as well as metabolites indicating infant renal development from birth (e.g. myo-inositol, 1-*N*-methylnicotinamide).

Overall, these investigations have shown that an understanding of the sources of variation in biofluid metabolite profiles are essential for interpretation of data acquired during normal infant development.

DECLARATION OF ORIGINALITY

I certify that this thesis and the research to which is accounted for here are the product of my own work. Any contributions or quotations from work of other individuals or published literature have been fully accredited in the accordance with the standard referencing practices of the discipline.

DISCLOSURE STATEMENT

Detection limits and urine contamination analysis was performed by the author; I would like to recognise that Nancy Georgakopoulou aided with the preparation of samples and Dr Michael Kyriakides for his help in metabolic identification from nappy contamination in infant urine (Chapter 3). All samples from the Norwegian Microbial (NoMic) study were prepared, analysed using ^{1}H NMR and multivariate data analysis was performed by the author. Dr Arnaud Wolfer assisted with time series analysis performed on this data set using his established algorithm (Chapter 4). Finally, urine and faecal sample data from the term vs preterm study was obtained using different analytical platforms, including ^{1}H NMR and LC-MS (Chapter 5). I would like to acknowledge the guidance from Dr Maria Romero-Gomez, Dr Manuja Kaluarachchi, and Dr Alex Pechlivanis in their advice during the analysis of these samples using LC-MS for untargeted analysis and targeted bile acid analysis.

COPYRIGHT DECLARATION

ACKNOWLEDGEMENTS

I would firstly like to thank my supervisors, Prof Jeremy Nicholson, Prof Neena Modi and Prof Julian Marchesi for this opportunity as well as their scientific input and direction over the last few years. I would especially like to thank Prof Elaine Holmes and Dr Anisha Wijeyesekera for their invaluable guidance, encouragement and support during my time at Computational and Systems Medicine (CSM); I will be forever grateful for this.

I'd like to acknowledge the STRATiGRAD funders of my PhD for giving me the opportunity to do this research at Imperial College London. I am also grateful for the members of CSM who contributed to the success of this project: Dr Maria Romero-Gomez, Dr Manuja Kaluarachchi, Dr Alex Pechlivanis and Dr Konstantina Spagou for their time and mass spectrometry expertise; Dr Michael Kyriakides for his help running NMR experiments and encouragement, Arnaud Wolfer for his time series analysis knowledge and Nancy Georgakopoulou for her help with nappy urine prep (and co-organising the postgraduate symposium!). Many thanks to my co-workers at Neonatal Medicine, especially Dr Matthew Hyde and Dr Nick Andreas for their invaluable help in further understanding this subject and making me feel welcome.

I would like to thank all my fellow students at CSM, especially the 1st year STRATiGRAD - Adam, Arnaud, Goncalo, Richie, Tom and Torben - who have been an enormous help over the last few years; not only with PhD related but also for their encouragement and valued friendship along the way. Also, many thanks to Isobelle for the breaks and walks around the block during write up.

On a personal note, I'd like to thank my friends, for their constant support, comforting words and much needed distraction over the last few years. An especially big thanks to my wonderful friends, Emma, Alessia, Tabitha, Rose, Caz, Nina and Beth for their continuous friendship and being my personal cheerleaders, especially in the last few months. I would also like to thank Gill for all her patience, support and encouragement.

Finally, I wish to thank my family for their love and encouragement, and always having full confidence in my abilities; my mum for being my rock, my dad for sharing his love of science and my brothers who I know will always have my back. I would also like to dedicate this work to my nana, who we miss every day and would have loved to see me finish.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVIATIONS

1D – One Dimensional
^{1}H – Proton
2D – Two Dimensional
3-HP- 3-hydroxyproline
3'-FSL – 3'fucosyllactose,
BCAA - Branched Chain Amino Acid
BM – Breast Milk
BMI - Body Mass Index
BW - Body Weight
BWG - Body Weight Gain
CE - Capillary Electrophoresis
CI - Chemical Ionisation
COSY – Correlation spectroscopy
CPMG - Carr-Purcell-Meilboom-Gill
CS – Caesarean section
DMA- Dimethylamine
DMG – Dimethylglycine
DNA – Deoxyribonucleic Acid
EI - Electronic Ionisation
EN – Enteral Nutrition
ESI - Electrospray Ionisation
FID - Free Induction Decay
FM – Formula Milk
GA – Gestational Age
GABA – γ-aminobutyric acid
GC - Gas Chromatography
HILIC - Hydrophilic Interaction Liquid Chromatography
HMO - Human Milk Oligosaccharide
JRES - J-resolved
MoD – Mode of Delivery
MS – Mass Spectrometry
NEC – Necrotizing Enterocolitis
NMN – 1-*N*-methylnicotinamide
NMR – Nuclear Magnetic Resonance
OPLS – Orthogonal Partial Least Squares Analysis
OPLS-DA – Orthogonal Partial Least Squares Analysis – Discriminate Analysis
OTU – Operational Taxonomic Unit
PBS – phosphate buffered saline
PC – Principal Component
PCA – Principal Component Analysis
PCR – Polymerase Chain Reaction
PLS – Partial Least Squares Analysis
PN – Parenteral Nutrition
PPROM – preterm premature rupture of membrane
QC – quality control
RNA – ribonucleic acid
SCFA – short-chain fatty acid
SD – standard deviation
STOCSY - Statistical Total Correlation SpectroscopY
TOF – Time of Flight
TMAO – Trimethylamine-*N*-oxide
TPN – Total Parenteral Nutrition
TSP – Trimethylsilyl Propionic Acid
VD – Vaginal Delivery

CHAPTER 1

INTRODUCTION

1.1. INTRODUCTION

Early life environmental exposures have been proven to impact health in later life, including mode of delivery, nutrition and the use of antibiotics during the neonatal period. These experiences have also been associated with disease risk; including increased possibility of developing specific allergies, asthma and even potentially metabolic syndrome in later life (Arrieta et al., 2015; Despres et al., 2006). The study of infant development has major implications for understanding the aetiology and risk of diseases, including metabolic syndrome and cardiovascular disease, in later life.

Metabolic profiling can be used to provide a baseline for knowledge and understanding of infant development and characterise the physiological or pathological processes through childhood and later life. There have been many more adult studies using metabolic profiling to further understand the biochemical composition of biofluids compared to neonatal and infant studies; this is especially true in regards to the normal metabolic trajectory of the healthy infant.

Human faecal metabolome studies is an area of research still very much in its infancy, in not only baby studies but adults too. Metabolic profiling of faeces can provide information on human health

but also the influence of diet and the gut microbiome (Saric et al., 2007; Jacobs et al., 2008). Both nutrition and the gut microbiome, which are intrinsically linked, in early life play a vital role in infant development (Chow et al., 2014). Further knowledge of the functionality of the gut microbiome on human development health would have a huge impact on further understanding this beneficial relationship.

The overall aim of this PhD thesis is to investigate the urinary and faecal metabolic profiles from full term and preterm infants, as well as characterise the developing infant faecal metabolome for the first time. This will generate new biochemical information on infant development in relation to gestational age as well as early life experiences including mode of delivery and nutrition.

1.2. PRETERM INFANTS

Gestational age at birth has been proven to have a large impact on the metabolic profile, as well as the composition of gut microbiome. At 40 weeks of gestation the infant is regarded as full term and ready to be born, but many infants born after 37-weeks of gestation are regarded as term as well. During the neonatal period, infants are rapidly adjusting to life outside of the womb; maturing and developing physically, immunologically, and neurologically (Carlo, 2016, Malina et al., 2004).

Preterm infants (<37 weeks of gestation) can be classified as early (before 32 weeks of gestation) and late (32-36 weeks of gestation) preterm. These can be further categorised into four different groups; extreme (GA ≤27 weeks), severe (GA 28-31 weeks), moderate (32-33 weeks) and near-term (34-36 weeks) (Moutquin et al., 2003; Wen et al., 2004; Tucker et al., 1991); these categories are important as lower gestational age is associated with increased adverse conditions; including mortality (Frey et al., 2016; Blencowe et al., 2012). In the UK, just over 1% of births are to infants born before 31 weeks of gestation, whereas this statistic rises considerably to 11% in developing countries (Goldenberg et al., 2008). Most of these births (about 8%) occur in the near term stage of preterm birth (34-36 weeks of gestation) period (Martin et al., 2013; Frey et al., 2016).

Birth weight of the infant can also give a better indication of how well the baby is developed as gestational age is difficult to calculate accurately. The birthweight of the infant generally indicates how well-nourished the infant is during foetal development and can be an important predictor of health in later life. This is especially true for brain development and subsequent IQ in later life (Grantham-McGregor et al., 1998; Walhould et al., 2012). The World Health Organisation (Edmond and Bahl, 2006) have defined low birth weight infants to be born as less than 2,500 grams. From their data, 15.5% of births worldwide are of infants of a low birth weight (WHO, 2006) and, as recorded in 2014, approximately 7% of births in the UK are to infants of low birth weight (Norman et al., 2005; ONS, 2015). Low birth weight (LBW) can arise from both preterm birth or fetal growth restriction, meaning that LBW doesn't necessarily mean that infants are preterm (Thomas et al., 2012). Knowledge of both the gestational age and birth weight is valuable in learning about the potential outcomes of infants (Simmons et al., 2010).

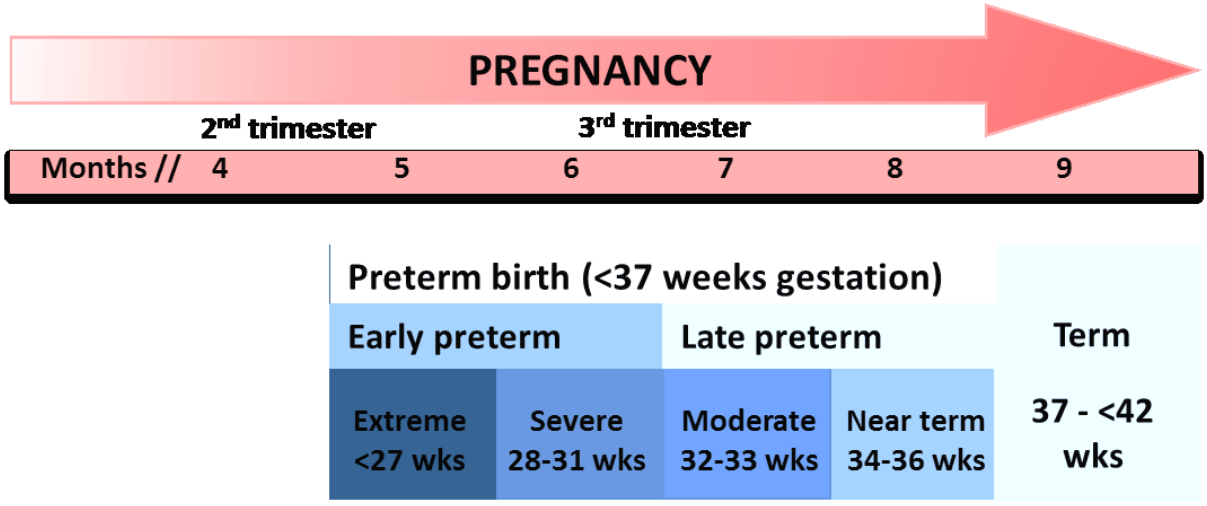


Figure 1-1: Schematic to represent the categorisation of term and preterm birth in relation to gestational weeks (wks) and pregnancy.

Prematurity is the leading cause of death in babies during the neonatal period in developed countries and there is much still unknown about premature birth. (Howson et al., 2013; Blencowe et al., 2012). While advances in neonatal medicine have seen an increase in the survival of babies born preterm since the 1990s; this has resulted in a significant burden to health and education services, with increased risk for several childhood and later life diseases (Behrman *et al.,* 2007; Blencowe et al., 2012). The aetiology of the most prevalent disorders as a result of preterm birth remains unclear. Preterm birth outcomes are more favourable with increasing gestational age, and birth weight is also a contributing factor (Tucker *et al.,* 2004). The main issue with preterm birth is that there is currently no method to predict the adverse outcomes of the event.

1.2.1. CONTRIBUTIONS AND TYPES OF PRETERM BIRTH

There are a number of different factors which can increase the associated risk of having a preterm birth. This includes maternal age, maternal lifestyle (including smoking and BMI), in-vitro fertilization, low-socioeconomic status, genetic influences as well as prior pregnancy history; this list is not exhaustive (Simmons et al., 2010; Frey et al., 2016; Moutquin, 2003; Goldenberg et al., 2008).

There are three main conditions which lead to preterm birth: spontaneous or idiopathic preterm birth, preterm premature rupture of membranes (PPROM) and medically indicated due to maternal or fetal difficulties (Moutquin, 2003). Averagely, medically indicated preterm births accounted for 25-35% of preterm birth, 45-50% were caused due to spontaneous births and 25-30% were caused by PPROM (Tucker et al., 1991; Goldenberg et al., 2008; Simmons et al., 2010). It is believed that over 40% of preterm births are caused by inter-uterine infections and that most early preterm births (less than 32 weeks of gestation) are also caused by this (Goldenberg et al., 2000; Carroll et al 1996; Goldenberg et al., 2008).

1.2.2. NUTRITION IN PRETERM NEONATES

Full-term infants are normally fed breast or formula milk orally, referred to as enteral nutrition. However, this is often not the case for infants born preterm, as they have different abilities depending on the gestational age at which they are born. This includes the ability to feed properly and what nutrients they need to develop outside the womb. Influence of the diet on the metabolism of the infant using metabolic profiling has not been investigated in depth yet. However, investigations using metabolic profiling has the potential to further understand the needs of the infants nutritionally.

They may need intravenous supplementation, called parenteral nutrition (PN) as well as enteral feeding. Furthermore, total parenteral nutrition (TPN) means that nutrition is not taken orally at all and only intravenously.

Neonatal nutrition for preterm infants aims to make sure that infants gain body weight (and ideally the same body development of a foetus of the same gestational age) and therefore, obtain the same functional outcomes that would be expected from an infant born full term (Brennan et al., 2016). Preterm infants are also more predisposed to postnatal weight loss and for being deficient in nutrients (Lapillonne et al., 2013).

1.2.2.1. Enteral nutrition

The different types of enteral nutrition include mother or donor human breast milk and formula feed for preterm infants at different stages of development (Brennan et al., 2016).

Human breast milk contains oligosaccharides which are complex carbohydrates. These are undigested by humans but utilised by the developing gut microbiome; this will be explained further later in this chapter. For many mothers, the composition of their breast milk is unique and recently it has been discovered that there is a higher concentration of oligosaccharides found in the mother's

milk from a preterm birth compared to full term (Bertino et al., 2013). Preterm infants which were fed mother's breast milk were half as likely to go on to develop nectrotising enterocolitis (NEC) or late on-set sepsis compared to being fed by donor breast milk or formula feed (Schanler et al., 2005; Cristofalo et al., 2013).

Human breast milk from the mother is recommended for any infant however this is not always available, especially from mothers with preterm infants (Jones et al., 2007; Radmacher et al., 2016). Donor human breast milk can be made available for preterm infants when their mother's milk is not obtainable (Bertino et al., 2013; Arslanoglu et al., 2013). This milk is usually pasteurised, removing any beneficial bacteria but also inactivating potentially pathogenic bacterial or viral agents (Arslanoglu et al., 2012). If no human milk is available, preterm formulas have been developed to ensure the nutritional needs of preterm infants are met to support growth. This formula needs to contain the correct composition and balance of lipids, carbohydrates and amino acids that are necessary to sustain this healthy growth in preterm infants (Hay et al., 2016).

Enteral feeding is important to start as soon as possible for any infant as it will decrease the risk of developing necrotising enterocolitis (NEC) as well as hypoglycaemia. However, due to the associated co-morbidities of prematurity, mainly abdominal abnormalities, enteral feeding may be delayed for preterm infants from birth compared to their full term counterparts (Patel et al., 2016).

1.2.2.2. Parenteral Nutrition

Even though breast milk is the recommended feed of choice, human milk is seen as insufficient and too variable in composition to provide preterm infants with all the nutrients they need, especially proteins and minerals (Wojcik et al., 2009; Radmacher et al., 2016). Due to their developmental immaturity the majority of preterm infants receive parenteral nutrition (PN) as well. This vital time in development means that not only is the infant needing additional nutrients for day to day growth but also to facilitate catch up growth due to prematurity.

This type of nutrition is administered intravenously, and is mainly for infants born before 34 weeks of gestation, as this is when motor skills needed for sucking, and therefore feeding, are developed. The time which the infant is spent on parenteral nutrition depends on gestational age, body weight and morbidities associated with prematurity. Preterm infants have low body stores which includes fat stored vitamins which they would have received through the placenta towards the end of gestation. Therefore, it has been recommended that these infants should be provided with water and fat soluble vitamins in PN feeds (Ho et al., 2015).

1.2.2.3. Total Parenteral Nutrition

Severe preterm infants and very ill babies have total parenteral nutrition (TPN) initiated and administered as soon as possible (Patel et al., 2016). As this type of nutrition is administered intravenously, these infants have no oral feeding experience and this could have a detrimental affect their body development including the lack of gastrin production (Ho et al., 2015). Furthermore, prolonged use of TPN has been found to increase body fat over time in different species (including piglets) (Meyers et al., 2013). This can lead to the higher risk of obesity, hypertension and insulin resistance later in life. These disorders have been recognised in adult metabonomic studies to have a perturbed metabolic profile compared to healthy subjects. By investigating the metabolic profile in early life, biochemical information may be available to predict future outcomes from this.

1.2.3. DEVELOPMENTAL DISORDERS ASSOCIATED WITH PREMATURITY

1.2.3.1. Early in life

During the neonatal period there is a significant risk of neonatal mortality and morbidity in preterm infants, with infants born before 32 weeks of gestation being at the highest risk (Thomas et al., 2012). Preterm infants are usually treated in a neonatal unit until the infant meets a set of criteria before discharged from hospital. These criteria determine the functional physiological maturity of the infant rather than relying on weight gain as a measure. These functions include:

thermoregulation, control of breathing, cardiac and respiratory stability as well as capable feeding skills (Jefferies et al., 2014).

Preterm infants are generally born immature as organ systems develop at different times of the gestational period. For example, a foetus' lungs in the womb are filled with liquid to help them grow and develop. At 36 weeks of gestation, the lungs are fully developed and during birth this liquid is absorbed so that the baby can start to breathe air once outside of the womb (Kotecha, 2000). Preterm infants born before this time have a higher risk of developing breathing problems as their lungs are not yet fully matured and may therefore need to be on a ventilator. Infants born before 34 weeks gestation are regularly administered surfactants, a complex mixture of phospholipids and proteins which coat the alveoli to improve the tension around it to keep it open; therefore this helps with oxygen intake (Nkadi et al., 2009).

As an extreme preterm infant, there is a much higher risk of having hypothermia, hypoglycaemia, a higher risk of low blood pressure as well as being susceptible to infections. The risk of these conditions reduce as the gestational age of the infant increases. Near to term infants, who are not likely to have similar breathing problems as their lungs are much more developed, however, still have some risk of blood sugar levels and blood pressure being too low, as well as potential feeding difficulties.

Microbial dysbiosis has been linked to several functional gut disorders, including inflammatory bowel disease, stomach cancer, obesity and necrotising enterocolitis (NEC) (Turroni *et al.,* 2012). Preterm infants are at a greater risk of developing NEC which can be fatal if not recognised early. NEC is a gastrointestinal disorder that affects primarily preterm babies and is marked by overgrowth of bacteria in the intestinal wall, which can lead to necrosis of the affected region of the intestine (Wang Y *et al.,* 2009). When NEC is first suspected, the infant is usually subject to TPN or fasting all together. The second recommended treatment for NEC is to initiate a short course of antibiotics, with some extreme cases requiring surgery to remove a diseased portion of the intestine (Wang *et*

al., 2009). All of these interventions results in the disruption of the microbiome and could cause further dysbiosis. Therefore, many studies are analysing the role of the gut microbiome in the development of NEC to help understand the cause and prevent it from occurring.

1.2.3.2. Later in life

Preterm and LBW babies who survive the neonatal period face the possibility of a lifetime of different disorders and disabilities (Howson et al., 2013). Due to their metabolic, anatomical and functional immaturity, these infants are associated with a higher risk of wide-ranging and long-term medical consequences. These include cardiovascular and metabolic disorders, abdominal obesity later in life and neurodevelopmental disorders (Moster, *et al.,* 2008).

Preterm birth may predispose children to the development of the symptom complex referred to as the metabolic syndrome. Metabolic syndrome starts with early insulin resistance, with the consequent development of type II diabetes, abdominal obesity, and hypertension (Abitbol *et al.,* 2012). Ex-preterm adults have higher plasma insulin, triglyceride, and cholesterol, compared to lower high-density lipoprotein cholesterol later in life (Irving *et al.,* 2000). Cardiovascular disorders are also very common and varied among preterm infants, including an association with a higher blood pressure when measured as young adults (de Jong *et al.,* 2012).

Preterm infants have low levels of body fat at birth as the third trimester is when the foetus will gain the main fat reserves (Thomas et al., 2012). Therefore, catch up growth using high fat nutrition is common but this also contributes to an increased adiposity (Brennan et al., 2016). Uthaya et al., (2005) found that preterm infants had increased abdominal adipose tissue and less subcutaneous adipose tissue compared to term infants, once they had reached term equivalent age. At adulthood, this increase in adiposity in preterm born individuals can still be seen (Thomas et al., 2011).

LBW and preterm babies are also associated with the development of insulin resistance and diabetes later in life. According to many studies in the past, not many very preterm infants would have

survived the neonatal period so could not be included in studies when researching preterm infants as adults (Thomas et al., 2012).

By 33 weeks of gestation, the nervous system of the foetus and the brain are fully developed. Many preterm infants develop cognitive and motor disabilities, including Cerebral Palsy and Autism Spectrum Disorders (ASD). It has previously been reported that up to 21-25% of preterm infants go on to develop ASD (Johnson, *et al.,* 2010). This is a much higher prevalence compared to the general population; 0.2-0.7%. However, other studies have suggested a 2 or 3-fold increase risk of ASD in preterm infants, hence the true risk is still unknown (Stephens *et al.,* 2012).

1.3. THE MUTUALISTIC RELATIONSHIP OF THE GUT MICROBIOME

The human gut consists of the stomach, small intestine and large intestine. These organs are responsible for the breakdown of food including carbohydrates, proteins and fats, which provide us with energy. The intestinal microbiota can also be described as another 'organ' (Eckburg *et al.,* 2005) as the composition of bacteria can provide our bodies with essential nutrients through homeostatic symbiosis (Leser and Molbak*,* 2009). We provide the bacteria with a stable environment, whereas the bacterium can provide us with nutrients such as, vitamin B12, which we cannot get from food.

In the human gut, there is thought to be over 10^{14} microorganisms, equivalent to about 10 times more bacterial cells than human cells in our bodies and most of which live within the colon. The diversity of these microbes is within a large range and intensely under investigation. This complex community of bacteria and some archaea, fungi and protozoa can help to keep our guts healthy by protecting us from the overgrowth of pathogenic bacteria (Gill *et al.,* 2006).

Each part of the GI tract differs greatly in function and environment, especially referring to pH levels, and this affects the species which live there. However, it is known that over 99% of these bacteria are from the bacterial divisions; *Bacteroides* and *Firmicutes* (Gill *et al.,* 2006). It is known that

Bifidobacterium and *Lactobacillus* species are the most common genera in the human colon and are important in promoting the host's health (Kim *et al.,* 2004).

Many of these bacteria have not been characterised and are uncultivable, which makes them very difficult to study and, therefore, many metagenomic studies have been used to explore this complex microbial community (Kuroawa *et al.,* 2007). The gut microbiota was first investigated through anaerobic culture techniques. Although, they did provide pioneering research about this microbial community, these techniques were very time consuming and they could only identify easily cultivable microorganisms. It was difficult to produce the right environment which resembled the gut for a more reliable indication of this environment microbiota composition. By using metagenomic analysis to isolate and clone DNA, the insight into the gut microbiota has improved our understanding of this uncultured world (Handelsman, 2004).

The human gut does have defences against the colonisation of microbes, which protect against pathogens and overgrowth. The most effective defence is the low pH content of the stomach that is harmful to most bacteria, which enter through the consumption of food. However, the bacterium *Helicobacter pylori* are able to survive in the stomach by producing the enzyme urease, which breaks down the acid around it and will colonise in the stomach's mucus (Leser and Molbak*,* 2009). This bacterium is commonly known to cause peptic ulcers, which can lead to cancer. The small intestine contains a much larger amount of bacteria than the stomach, which eventually leads to the large intestine, where there is the highest concentration of bacteria. In the small intestine there are antimicrobial systems that keep control of the growth of the bacteria. This includes the fact that, as material travels through the gut, it dislodges some of the bacteria attached to the epithelial cells. It also includes the secretion of antimicrobial peptides, proteolytic enzymes and bile (Ridlon *et al.,* 2006). If these methods stop working then the bacteria will overgrow and we would have to compete for our nutrients much more, resulting in non-symbiotic relationship and possible infection (Inagaki *et al.,* 2006).

The gut virobiota is another area of research which has expanded recently; however, there are also many challenges that face these types of studies and is many aspects are still largely unclear (Duerkop & Hooper, 2013). It is believed that many of the virus which reside in our guts are bacteriophages (viruses that infect and replicate inside bacterial cells), which can in effect influence the composition of the bacterial community also. However, both eukaryotic and plant viruses (from the diet) have been found to be present as well (Handley et al., 2012). Investigations into the infant gut virobiota found that, similar to the bacterial community, there is a low diversity at birth and early life. It was also found that the most abundant viral species did not originally come from breast milk, suggesting another source may seed the viral community in early life (Breitbart et al., 2008).

1.3.1. FUNCTIONS AND METABOLIC ACTIVITIES OF THE GUT MICROBIOME

The function of the gut microbiome can be very complex due to the diversity of enzymes that the microbes utilise. These enzymes are usually different to the ones that we produce and contribute to many different metabolic pathways.

Studies on germ free (GF) mice also emphasises the importance of the gut microbiota on development; physically, biochemically and immunologically. Mice without a gut microbiome have a larger caecum, decreased intestinal surface area and many of their organs are reduced in size; including the liver, heart and adrenal glands (Berg et al., 1996). Furthermore, the effectiveness of the immune system as well as the inflammatory responses are also greatly reduced in the GF animal (Berg et al., 1996). Therefore, the presence and mutualistic relationship of commensal bacteria is thought to be essential in maintaining the physiology of the digestive tract as well as other organs in the body.

For bacteria in the colon, most of the available dietary components are the insoluble particles which the human body cannot break down. These undigested carbohydrates (e.g. starches, cellulose, and some oligosaccharides) are mainly fermented into short chain fatty acids (SCFAs) and gases (e.g.

hydrogen, carbon dioxide and methane). SCFAs are very useful as they are a local energy source for the colonic epithelial cells (especially butyrate) as well as some becoming readily available for the rest of the body (especially acetate) (Flint et al., 2012). Fermentation from bacterial species of high protein diets increases the concentration of branched chain amino acids (BCAA) into branched chain fatty acids (Russell et al., 2011). These metabolites have been associated with hormone receptor control and therefore involved in appetite control. This shows a potential link between microbial activity and a contribution to the need to eat (Sleeth et al., 2010). Colonic microorganisms are also known to be involved in the synthesis of vitamins (B & K) from the production of SCFAs and the absorption of iron and magnesium (Guarner and Malagelada, 2003).

Choline is an essential nutrient found in many different foods including eggs, soy beans and milk. It is mainly used to make phosphocholine for the construction and stability of cell membranes in the body. The gut microbiome utilises this metabolite and catabolises it into trimethylamine (TMA); TMA is then absorbed by the body and further metabolised into trimethylamine-n-oxide (TMAO) in the liver. An increase in urinary levels of TMAO has been reported after the consumption of fish within 24 hours (O'Gorman et al., 2013). Lloyd et al. (2011) investigated the urine from healthy participants after they had consumed specific foods; one of these foods was salmon. Urine was taken at different time points after the consumption of the foods, analysed using mass spectrometry and metabolites were identified from the result of eating the specific foods. It was clear, from this investigation, that TMAO, anserine and 1-methylhistidine was highly associated with the consumption of fish (Lloyd et al 2011). Equally, high concentrations of TMAO in human plasma has also been implication in the potential risk for liver and cardiovascular disease (Nicholson et al., 2012).

Bile acids are cholesterol steroids that are conjugated with glycine or taurine and are effective lipid detergents. Some gut microbes can deconjugate and metabolise bile acids to make them harmless and even use them as a source of carbon, nitrogen and energy sources (Begley et al., 2005). In

addition, the microbiota actively regulates fat storage by promoting lipid absorption through the chemical modulation of bile acids.

Another link to health which is currently under intense investigation, is that SCFAs, including butyrate and propionate, have anti-inflammatory properties and therefore anti-cancer properties (Flint et al., 2012). This relationship between the microbiota and cancer development is still unclear. Furthermore, SCFAs can interact with G protein-coupled receptors (GPCRs) to affect insulin sensitivity in adipocytes and peripheral organs, thus regulating energy metabolism (Backhed et al., 2004).

The composition of the gut microbiota is dramatically influenced by the host's diet and how the gut microbiome is altered by obesity. Jumpertz et al. (2011) investigated the change in the composition of the gut microbiome in lean and obese individuals when they had their diets changed and nutrient load is changed. They concluded that the amount of nutrient that individuals consume greatly influences the composition of the gut microbiome and an increase in the community of *Firmicutes* was associated with an increase in calorie intake in lean individuals (Jumpertz et al., 2011). Other studies have also revealed that patients suffering from obesity have a higher number of *Firmicutes* residing in their gut (Le Chatelier et al., 2013; Ley et al., 2006). These bacteria are more adapted to breaking down carbohydrates, which our body cannot breakdown normally and from which we cannot gain nutrients. However, when obese individuals are put on a diet containing less fat over a course of a year, it was found that the *Firmicutes* numbers slowly decreased and *Bacteroides* increased in abundance (Frank and Race, 2008). Further studies have also revealed that a high fat diet can also alter the composition of the gut microbiome; with a lower abundance of bacteria including *Bacteroides* and *Bifidiobacterium* (Nava et al., 2012). These studies show the influence of the microbiome on potential energy harvest from our diets (depending on what is available).

1.3.2. DEVELOPMENT FROM BIRTH

The gut is quickly colonised from birth by microorganisms; at first, the newborn gastrointestinal tract is an aerobic environment where only facultative anaerobes are able to live (Arrieta et al., 2015). Over time, larger numbers of anaerobic species establish in the infant gut and drive out the facultative bacteria through competition for nutrients. Early infant microbiome has a dominance of *Bifidiobacterium*, *Clostridium* and *Bacteroides*; this community develops through the first few months of life (Matamoros et al., 2012, Collado et al., 2012, Nicholson *et al.,* 2012). It is believed that gestational age, the mode of delivery, nutrition (maternal or formula milk), clinical interventions (e.g antibiotics), environmental exposures as well as maternal microbiota can affect the composition of the early gut microbiome.

However, it has recently been postulated that the foetus is not exclusively sterile as previously thought; bacteria has been isolated in the umbilical cord blood and the meconium of healthy full term babies (Jimenez *et al.,* 2007). Furthermore, bacteria have also been isolated in the amniotic fluid and placentas of full term infants, where it would have usually been associated with preterm birth (Staokari et al., 2008; Aagaard et al., 2014). This suggests that gut colonisation starts before birth and could affect the *in utero* development in a number of different ways; whether this is beneficially or detrimentally, is still unknown.

Eventually, the microbiota is established and stabilised by the end of the third year of an infant's life. This composition is unique to the individual and will not change dramatically throughout life, although it is continuously fluctuating (Palmer *et al.,* 2007). There is a reported familial similarity, and it is thought that genetic makeup, as well as environmental factors are contributors to the determination of the composition (Leser and Molbak, 2009). It has also been reported that with an increase in age, the diversity of bacterial species in the gut decreases, meaning the composition is very variable throughout life (Ottman et al., 2012).

1.3.3. MODE OF DELIVERY

The mode of delivery, vaginally delivered versus caesarean section, has a strong influence on the early gut microbiome colonisation in infants. It has been well documented that infants who have been delivered through the birth canal, have a bacterial community which is similar to that of a typical vaginal community and the mother's gut microbiome; this includes *Lactobacilli, Prevotella* and *Sneathia* (Dominguez-Bello et al., 2010). Furthermore, vaginally delivered infants are primarily colonised by different *Bifidiobacterium* species within the first couple of days, the most prevalent species being *B.catenulatum* group and *B.longum* (Biasucci et al., 2008; Biasucci et al., 2010).

Caesarean section (CS) infants are influenced by the exposure to the mother's skin and the clinical environment they first experience. Therefore, these infants are associated to different bacterial species compared to vaginally delivered; including *Staphylococcus*, *Corynebacterium* and *Propionibacterium* (Dominguez-Bello et al., 2010).

Some studies have found that infants born by caesarean section have a lower diversity of bacterium and so their acquisition of *Bacteriodes*, *Bifidobacteria* and *Escherichia coli* is delayed compared to infants born by vaginal delivery (Azad et al., 2013; Dominguez-Bello et al., 2010). However, there is still much to learn, as there are many contradictory results currently in the literature, especially in terms of diversity and presence of pathogenic bacterium (Rotimi et al.,. 1985, Hallstrom et al., 2004; Rutayisire et al., 2016).

Furthermore, these infants are more likely to be colonised by *Clostridium difficile* compared to vaginally delivered infants (Penders et al., 2006). This bacterium is a toxin producing bacteria, which mostly causes diarrhoea and food poisoning (Schutze et al., 2013). A study of the gut microbiome composition related to mode of delivery in children at age 7 showed that *Clostridium* species were higher in children born by CS (Salimen et al., 2004). However, other studies found no significant associated between mode of delivery and the prevalence of *Clostridium difficile* (Azad et al., 2013).

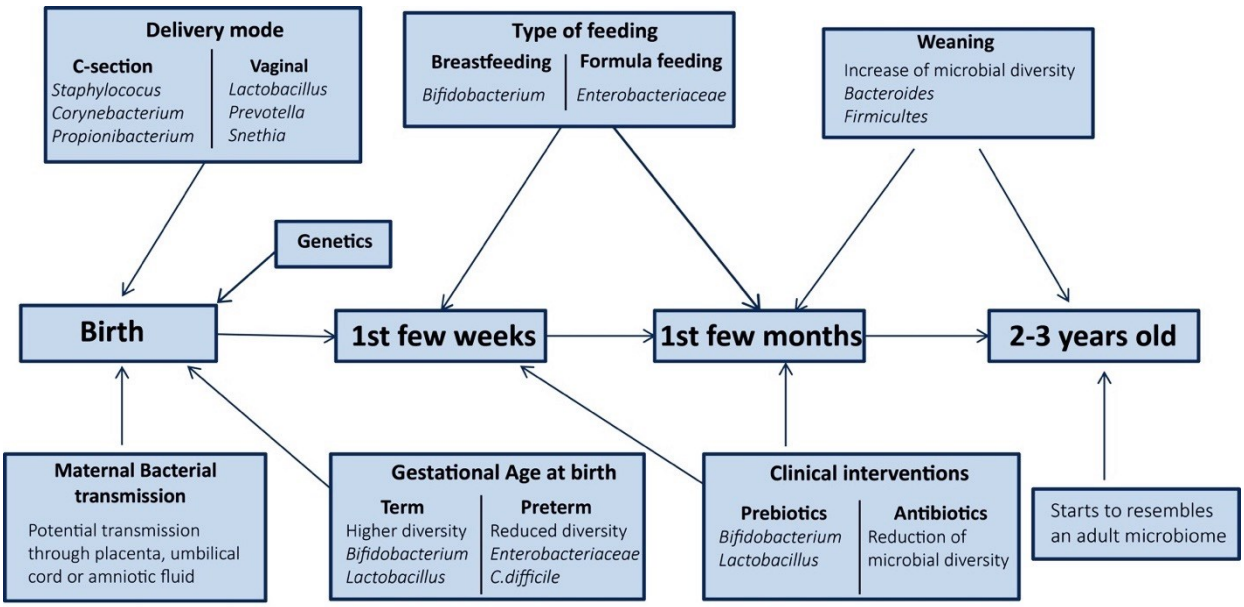


Figure 1- 2. Schematic to illustrate the influence of external factors on the developing gut microbiome from birth to infancy. Adapted from Matamoros et al., 2013. (Collado et al., 2012, Voreades et al., 2014, Rodriguez et al., 2013)

1.3.4. GESTATIONAL AGE

Preterm infants generally have a lower diversity of gut microbiota during the first few weeks of life compared to full term babies (Cong et al., 2016). Initially, preterm infants have difficulty digesting nutrients as some specialised cells are not fully functional, and this may affect the composition of the microbiota. Since preterm infants are generally underdeveloped, they may contribute to gut dysbiosis as a result of immature peristaltic ability, which is used to control the microbiota community (La Rosa et al., 2014). Furthermore, preterm infants are subject to more antibiotic treatments and different early nutrition than full term infants, and so it is difficult to form conclusions from these studies.

Recently, Cong et al., (2016) discovered that preterm infants had a high proportion of Proteobacteria and there was a significant difference between male and female infant gut community. There are not many studies comparing preterm and term infants gut microbiome composition, due to the

different experiences that these two groups of infants go through in early life. On a species level however, it has been found that preterm infants have a higher diversity of facultative anaerobes and less strict anaerobes compared to full term infants (Arboleya et al., 2012; Stark et al., 1982). Stark et al., (1982) found that preterm infants had a delayed colonisation of *Bifidobacterium* species compared to term infants.

1.3.5. NUTRITIONAL IMPACT ON THE INFANT GUT MICROBIOME COMPOSITION

It is universally accepted that feeding by breast milk for the first 6 months is very beneficial and even to 1 year with the introduction of solid foods is ideal practice (Bertino et al., 2013). In the first few days after birth, a specific type of milk is produced called colostrum. This milk is low in fat and high in carbohydrates, proteins and antibodies; it also acts as a laxative to remove the meconium from the intestines.

Babies who are exclusively fed breast milk have a more stable, lower diversity bacterial composition compared with formula feed infants; breast fed infants have a much higher abundance of *Bifidobacterium* and *Lactobacillus* species (Penders et al., 2006; Matamoros et al., 2013; Harmsen et al., 2000; Roger et al., 2010). These infants also have other microbes present, including *Staphylococcus* and *Propionibacterium* which are transferred from the skin and nipple of the mother (Collado et al., 2012). Formula fed infants have a higher diversity of bacteria in their gut, mainly *Bacteroides*, *Clostridium* and *Enterobacteriaeae* species (Penders et al., 2006; Fallani et al., 2010; Gomez-Llorente et al., 2013).

Breast feeding is believed to be more functionally beneficial and better protected against infection of the GI, respiratory and urinary tracts as well as other diseases compared with those who are formula fed. Salminen and co-workers have attributed this effect, at least partly, to differences in microbial composition (Salminen et al., 2005). Furthermore, this beneficial protection can be attributed to many different aspects of breast milk as well including antimicrobial enzymes (e.g.

lysozyme) found in breast milk which helps to inhibit the growth of certain bacteria (Kau et al., 2011). Human milk oligosaccharides (HMO) are unique complex carbohydrates and cannot be digested by human intestinal enzymes. Therefore, they reach the large intestine and colon intact, where they can promote the proliferation of certain types of bacterial species (Arrieta et al., 2015). Breast milk also contains IgA antibodies which are similar to bacterial antigens to prime the immune system against common pathogens (Kelly et al., 2000).

Feeding with parenteral feed is sometimes quite necessary for preterm or very sick infants, however this can cause a lower diversity in the gut microbiome as well as a delayed colonisation (Blakey et al., 1982; Jacquot et al., 2011). These infants are also more associated to the bacterial species *Pseudomonas* and *E. coli*.

When the infant starts weaning, usually around six months, the gut microbiome changes again and starts to resemble an adult gut microbiome (Harmsen et al., 2000). In general, the *Bifidobacterium* community gradually decreases and a more complex composition establishes itself, with the main species being *Bacteroides*, *Clostridium* and *Eubacterium* (Roger et al., 2010). It is difficult to know exactly how the gut microbiome will change at this stage as it is very dependent on the more complex diet becoming available, supporting a more diverse bacterial community.

1.3.6. MEDICAL AND ANTIBIOTICS INFLUENCE

In neonatal intensive care units, antibiotic treatment is a common occurrence to reduce bacterial infections that could be fatal; this is especially true for infants born extremely premature. Antibiotic over use in this setting is a cause for concern as antibiotic induced microbial dysbiosis could affect early life development (Jernburg et al., 2010; Arrieta et al., 2015).

Overall, total count and diversity of the gut microbiome is decreased with antibiotic use (Koenig et al., 2011; Blakey et al., 1982; Butel et al., 2007; Tanaka et al., 2009). Broad spectrum antibiotics can

significantly disrupt the composition of the microbiome, potentially causing the susceptibility of pathogenic bacteria to colonise the gut (Collado et al., 2012; Arrieta et al., 2015).

Infants have a decreased colonisation of *Lactobaccillus* and *Bifidobacterium*; common beneficial bacterial species found in infants (Blakey et al., 1982; Penders et al., 2006; Butel et al., 2007). These infants are commonly colonised by *Enterobacteraceae*, *Klebsiella* and *Clostridium* which dominate the community; this is due to these microorganisms being highly resistant to broad-spectrum antibiotics (Alderberth et al., 2006; Penders et al., 2006; Tanaka et al., 2009; Collado et al., 2012).

Even though antibiotics have been proven to affect gut microbiome composition which could further affect health in later life, many studies are researching different types of antibiotics which can affect different bacterial species. Furthermore, these infants are extremely unwell, and therefore dysbiosis of the gut microbiome is probably occurring; this adds to the complex nature of the interaction, making it difficult to study.

1.3.7. DEVELOPMENT OF THE IMMUNE SYSTEM

The colonisation of the gut induces immune development through immune homeostasis as well as promoting the host-microbial mutualistic relationship (Geuking et al., 2014). Host-microbe interactions occur mainly along the mucosal surface and the gut microbiota has a vital role in immune system development. This includes both food and bacterial antigens from the environment which are harmless and allow regulation of inflammatory responses (Bischoff et al., 2011). In the large intestine, the mucosal layer is the main means by which bacterial overgrowth and penetration via the epithelial cell surface occurs. The mucosa is split into two layers, the outer layer is colonised with bacteria and the inner layer has antimicrobial lectins, which are expressed by epithelial cells, constantly being secreted (Hooper et al., 2012). This diverse community of microbes causes an abundance of foreign antigens interacting with both the mucosal immune system as well as the systemic immune system. This interaction shapes the development of the immune system and

causes a 'cross-talk' between the gut microbiota and the immune system (Nicholson et al., 2012). The interaction is still not fully understood as the immune system must find a balance between allowing the growth of bacteria and discouraging overgrowth.

Furthermore, breast milk contains antibodies from the mother which facilitates the adaptive immune system whilst the infant is fed. If the infant is formula fed, they rely on their own white blood cells to fight any pathogenic antigens. During weaning, the infant's immune system is further developed through the introduction of new foreign antigens, and their bodies rapidly produce their own antibodies (Hooper, 2004).

1.3.8. NEUROLOGICAL DEVELOPMENT ASSOCIATED TO THE GUT MICROBIOME

Sudo *et al.,* (2004) postulated that the gut microbiota has a role in the development of hypothalamic-pituitary-adrenal (HPA) function during the first weeks of life and can therefore have an impact on the enteric nervous system (ENS) (Sudo et al.,. 2004; Nicholson *et al.,* 2012). There has been increasing interest in the gut-brain axis and whether the gut microbiota and diet can affect brain development and function, including behavioural changes (Collins *et al.,* 2009).

For example, autistic spectrum disorders (ASD) are a group of developmental disabilities that are associated with impaired social skills, communicative skills and repetitive behaviours. (Parracho et al., 2005). Recent research has shown that ASD is linked not only to the neurological system but also to the dysregulation of the immune and digestive system. Some autistic children have a range of different gastrointestinal disorders which can be associated with an abnormal gut microbiota. This may be attributed to ASD children receiving antibiotic therapies during infancy, which has been documented to lead to gut dysbiosis (Adams et al., 2011). It has been hypothesised that the toxins produced by the gut microbiota may play a role in the etiology of ASD, in particular, lipopolysaccharides (LPS) (Gondalia et al., 2010).

1.3.9. GUT MICROBIOME INFLUENCE IN HEALTH LATER IN LIFE

Colonisation of the gastrointestinal tract by microbes is an essential process, as the interaction between host and microbiota have an important influence on human health and disease. Digestive function and health is dependent on the composition of the microbiota, which if disrupted can lead to 'dysbiosis'. This can lead to complex diseases including inflammatory bowel disease, celiac disease and colorectal cancer. It is becoming increasingly important that this relationship and the impact of molecular interactions in nutrition and digestive health is understood.

Each part of the GI tract differs greatly in function and environment, especially referring to pH levels, and this affects the species which live there. However, it is known that over 99% of these bacteria are from the bacterial divisions; *Bacteroides* and *Firmicutes* (Gill et al., 2006). This complex community of bacteria and some archaea, fungi and protozoa can help to keep our guts healthy by protecting us from the overgrowth of pathogenic bacteria. It is known that *Bifidobacterium* and *Lactobacillus* species are the most common genera in the human colon and are important in promoting the health of the host (Kim et al., 2004).

1.3.9.1. Obesity

This condition is an imbalance of higher intake of energy compared to the expenditure of energy. It is believed that certain exposures early in life can play an important role in the long term risk of obesity (Biro et al., 2010; Lifschitz, 2015). Studies on germ free mice have provided evidence for an association between gut microbiome and body weight (Let et al., 2005). Transfer of the gut microbiome from obese mice that are genetically prone to the disorder, to germ free mice, increased the fat deposits in these animals regardless of an increase in food intake (Ley et al., 2006).

Furthermore, studies in humans have revealed that patients suffering from obesity have a higher number of *Firmicutes* living in their gut (Ley et al., 2006). These bacteria are more adapted to breaking down food, which the body cannot breakdown normally and from which it cannot gain

nutrients. However, when these obese individuals were put on a reduced fat diet over the course of a year, it was found that the *Firmicutes* numbers slowly decreased and *Bacteroides* increased in abundance (Frank and Race, 2008). Therefore, it has been postulated that the bacterial community of the human gut is dramatically influenced by the host's diet.

1.4. OMICS TECHNOLOGIES

Metabolic phenotyping is the study of low molecular-weight endogenous metabolites present in biological fluids including urine, plasma and faecal water (Nicholson et al 2005). This analytical technique generates a unique metabolic "fingerprint" from biological samples, and thus is an invaluable tool in learning more about the complexities of diseases (Nicholson et al., 1999).

Metabolite profiles can be generated using different high resolution analytical platforms including nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). Global profiling using NMR and Ultra Performance Liquid Chromatography – Mass spectrometry (UPLC-MS), allows high throughput analysis of samples, identifying the most significant and highly concentrated metabolites by statistical analysis, with no *a priori* knowledge. Targeted profiling requires knowledge of specific metabolites of interest and uses more sensitive analytical techniques including liquid chromatography (LC) and gas chromatography (GC) hyphenated MS approaches, which are able to discern specific groups of molecules such as bile acids and short chain fatty acids (SCFAs). Furthermore, gut microbial associated metabolites can be detected in human biofluids, particularly in urine and faeces.

Combined 'omic' -approaches can be used to examine complex disease associations, and can be invaluable to investigate gut microbiota composition and the correlation to metabolic regulation. Metagenomic analysis can establish the composition of gut microbiota through surveying the 16S rRNA genes (Li et al., 2011). By combining metagenomics with metabonomics, it supports the identity of microbiome-based biomarkers of metabolic activity. This knowledge has the potential to

aide us in identifying new ways to manipulate the microbiome to optimise their function for individuals in health and disease.

1.4.1. METABONOMIC ANALYSIS IN INFANTS

The clinical potential of metabonomics in neonatology has become evident from the rapidly expanding literature reported related to this area, demonstrating how physiological foetal development or pathological conditions affect the perinatal and neonatal outcome (Martin et al., 2014; Moco et al., 2013).

The composition of urine mirrors many metabolic processes and therefore provides information on metabolic changes in different organs. Further understanding of this post-natal organ development will help indicate better care initiatives for infants or may help provide us with biomarkers for developmental disorders. This is especially true for preterm birth and development, as it is a significant and increasing burden on the healthcare system.

Metabonomics is a promising approach to investigate the interaction between our bodies and the gut microbiome, using metabolic profiling to capture endogenous metabolites as well as co-metabolites from microbial sources. These exogenous compounds, which originate from the diet and/or gut microbiota, can enter the host's system where they can be further processed in the liver and other tissues before being excreted in the urine. Therefore, urine is an ideal matrix in infant research as it is non-invasive and can be influenced by the diet and gut microbiota activity. Furthermore, consecutive samples help us to understand the dynamic changes from the rapidly developing body over time.

Additionally, human faeces are also an excellent source of biological information to interrelate the human microbiome and human metabolome. The metabolic profile of faecal water can provide a window for understanding the complex metabolic relationship between humans and the complex ecosystem residing in the gut. The normal metabolic trajectory of the healthy neonate and how the

development of the gut microbiome can influence the faecal metabolic profile, remains relatively unknown. The aim of this thesis is to further explore this area of research.

1.4.1.1. Urine Metabolic Profiling

Urine is usually the biofluid of choice for metabolic profiling analysis, as metabolites in the urine are known to have entered the body systemically. Additionally, neonatal urine is typically used to test for many different inborn errors of metabolism using different assays available (Constantinou et al., 2005).

Age is an important contributor to the metabolic profile with many studies looking at birth up until early childhood (Jackson et al., 2016). In one study, urine from infants collected at birth until age 12, identified several metabolites correlated to age including creatinine, creatine, glycine, betaine/TMAO, citrate, succinate, and acetone (Gu et al., 2009). Martin et al., (2014) also investigated the metabolic profiles of urine samples taken from the same infant at different time points. They discovered that the metabolic profile changed over time with an increase in two ketone bodies, 3-d-hydroxybutyrate and acetoacetate; these metabolites contribute to energy metabolism (Martin et al., 2014). Furthermore, authors discovered that these metabolites were present at different concentrations in breast fed versus formula fed infants. Recently, Chiu et al., (2016) also investigated the urinary metabolic profiles of healthy infants over time from six months to 4 years. They discovered that at 6 months of age, urinary excretion of betaine, trimethylamine-*N*-oxide, glutamine and glycine were higher compared to age 1; these samples had increased concentrations of creatine and creatinine by comparison (Chiu et al., 2016). Another group (Diaz et al., 2015) found metabolite changes in healthy new born infants dependant on gender and mode of delivery, these metabolites included betaine, 4-hydroxyphenylacetate and myo-inositol among others (Diaz et al., 2015).

Metabolic profiling has given a better insight into kidney development outside of the womb and the risk of renal disorders later in life. Even though the kidneys are sufficient to maintain homeostasis in the infant, its filtration function is not optimum (Drukker et al., 2002). Foxall et al., (1995), found that osmolytes, including taurine, myo-insitol and trimethylamine-n-oxide in the urine helps to protect medullary cells found in the kidney. Furthermore, they also found betaine (which is synthesised from choline in the diet) increases throughout the neonatal period (Brown et al., 1989; Foxall et al., 1995; Trump et al., 2006).

Gestational age has a strong effect on the urinary metabolic profile. In studies investigating preterm infants compared with full term infants, an altered urinary metabolome was discovered. Concentrations of urinary alanine, formate and citrate were found to be higher in preterm infants whilst creatinine, creatine and dimethylglycine were lower. This altered metabolome persisted until the preterm infants reached term-equivalent age and also when ex-preterm individuals were tested as young adults (Hyde *et al.,* 2010; Parkinson et al., 2013). Another study showed an inverse correlation of urinary hippurate with weight in preterm infants and, therefore, may be related to elevated adiposity later in life (Thomas *et al.,* 2011).

Furthermore, infants born preterm had lower concentrations of tryptophan, phenylalanine, malate, tyrosine, hydroxyl-butyrate, *N*-acetyl-glutamate, 1-*N*-methylnicotinamide, and proline than term born infants, but similar concentrations of betaine (Atzori et al., 2011; Trump et al., 2006). Preterm infants were shown to have higher concentrations of urinary bile acids, which was attributed to an overproduction or bile acid clearance (Yamato et al., 2001). Diaz et al., also found a range of metabolite differences in urine profiles from preterm infants when compared to infants with other disorders (e.g. respiratory distress, large for gestation age and malformation) (Diaz et al., 2016).

Another interesting research area of infant metabolic profiling is in relation to neurological disorders, including ASD, and the gut microbiome. It is known that increased levels of urinary indolyl-3-acryloylglycine (IAG) are seen in individuals with ASD, which is an indication of gut dysbiosis. It has

been theorised that an increased level of IAG causes a disruption in the gut membrane and increases the intestinal permeability (Shattock and Whiteley, 2002). This can then allow an unusual amount of undigested peptides or bacterial derived toxins to enter the blood stream and possibly cross the blood brain barrier (Wang *et al.,* 2009). Yap *et al.,* (2010), discovered that there was a significant difference in the metabolic composition of urine between ASD patients, their unaffected siblings and non-related controls. Urinary concentrations of hippurate and phenylacetylglutamine (PAG) were found to be lower in ASD children, as well as perturbations in urinary amino acid concentrations and metabolites linked to nicotinic acid metabolism (Yap *et al.,* 2010). Furthermore, several autistic children have different levels of *N*-methyltryptamine, L-arabinose and 4-cresyl sulphate which also adds evidence of gut abnormalities in these individuals (Altieri *et al.,* 2011).

1.4.1.2. Faecal Metabolic Profiling

Human faecal metabolic profiling is a relatively new research area compared to other biofluids, with many studies only utilising this biofluid in the last 10 years (Saric et al., 2008; Jacobs et al., 2008; Monleon et al., 2009). The application of faeces has shown to have great potential for non-invasive diagnosis of diseases and disorders. For example, Marchesi et al., (2007) showed that faecal metabolic profiling by ^{1}H NMR can be used to diagnose inflammatory bowel disease (IBD). Furthermore, faecal metabolic profiling has shown to help understand the interaction and influence of microbiome co-metabolites on human health (Le Gall et al., 2011). Therefore, faecal metabonomics is becoming a useful tool for defining interactions among host, microbial communities and nutritional interventions (Nicholson et al., 2005; Jacobs et al., 2008; Martin et al., 2010).

From birth, the gut microbiome development and interaction also plays a crucial role in later health, especially in terms of early nutrition (Chow et al., 2014). The effect of breast feeding versus formula feeding is a common topic of interest in infant development. Recently, studies based on faecal metabolic profiling has been utilised to further expand this knowledge. It was found that protein

fermentation by the gut microbiome was seen at a higher rate in formula fed infants compared to breast fed infants (Chow et al., 2014; Martin et al., 2014). This is thought to be due to the higher concentration of proteins in formula feed to make up for the lack of human milk oligosaccharides.

Little is known about the faecal metabolome in infants and how it changes from birth during such a rapidly developing period. Therefore, further investigation is needed into not only faecal metabolic profiling but infancy development.

1.5. SCOPE OF THIS THESIS: Develop a metabonomic strategy to further understand early life metabolism in developing infants

The first goal of this work was to optimise and standardise the use of infant biofluid samples for metabolic profiling. This included characterising any contaminants that can come from the collection materials used to collect, prepare and run infant biofluid samples by ^{1}H NMR and UPLC-MS, as well as investigating the detection limits of different biofluids (urine, plasma and faeces) using ^{1}H NMR. Furthermore, understanding the age variability of urine metabolic profiles was critical to interpreting results throughout this project (Jackson et al., 2016). Finally, this work produced and optimised a protocol for infant faecal water extraction for both ^{1}H NMR and UPLC-MS. As previously mentioned, faecal metabolic profiling in humans is a relatively new method and so an optimised preparation for infant samples was lacking from the literature or within the department at the start of this project.

After this investigation was completed, the developed faecal water extraction method was carried out on two different sample sets; a term vs preterm study and the Norwegian Microbial (NoMic) study. Faecal samples collected as part of a large non-clinical cohort study, called the Norwegian Microbial (NoMic) study, using ^{1}H NMR spectroscopy. Samples were collected from 524 infants at 4, 10, 120, 365 and 730 days postpartum, which for many metabolic profiling-based infant research studies, have not had this number of time points, or participant numbers on this scale. This longitudinal study facilitated the characterisation and understanding of the faecal metabolic profile

of developing infants using two different types of statistical analysis; including multivariate statistics and time series analysis.

Based on current literature, there are few investigations which have compared term and preterm infants using metabolic profiling; especially using more than one biofluid. Thus, in this investigation the aim was to better understand the early life metabolism in term and preterm born infants. This sample set consisted of urine and faecal samples from both term and preterm born infants; longitudinal samples were taken from birth until three months postpartum. An investigation comparing metabolic profiles between term and preterm infants was conducted at different stages of life: at birth, term equivalent age and three months postpartum. Utilising ^{1}H NMR, UPLC-MS for global metabolic profiling and metagenomics analysis, has given an in-depth understanding of the differences between these two groups from birth and through the first few months of life.

To date, this critical period of development and how it impacts future health of the child has not been fully understood nor investigated using this analytical approach. By using metabolic profiling to explore the functionality of the developing gut microbiome, the role of microbial molecules in host physiology and health will be better understood.

CHAPTER 2

ANALYTICAL METHODS

Metabolic profiling techniques are applied to a wide range of complex samples such as biofluids and tissues. Profiles derived from urine and faeces give insight into specific metabolic changes and provide indirect information about gut microbial activity as urine contains specific metabolites produced by host-microbial interactions. It has been estimated that there are approximately 2000 metabolites making up the human metabolome (Beecher, 2003). However 2651 compounds have been identified in human urine alone (Bouatra et al., 2013).

Metabolite profiles can be generated using different high resolution analytical tools including Nuclear Magnetic Resonance (NMR) spectroscopy and Mass Spectrometry (MS) coupled with Liquid Chromatography (LC) (Nicholson et al., 2012). The advantages of using ^{1}H NMR spectroscopy is that it is a non-destructive technique; therefore, it doesn't destroy nor change the sample and can be used for further analysis. Furthermore, this technique is robust and more reproducible compared to mass spectrometry. The disadvantage of using ^{1}H NMR is that it is not as sensitive as UPLC-MS. Mass spectrometry when coupled with liquid chromatography can be highly sensitive and selective for targeted metabolic analysis including bile acid and SCFA analysis. The disadvantage of mass

spectrometry is that it is destructive to the sample as well as harder identification of metabolites due to matrix effects. These techniques have advantages and disadvantages, but by utilising these complimentary techniques, the information recorded from as many metabolites as possible gives a more in depth picture of the metabolome.

2.1. NUCLEAR MAGNETIC RESONANCE

High resolution proton NMR spectroscopy is a quantitative and reproducible analytical method, which doesn't destroy the sample. It is used to detect a wide range of low molecular weight compounds (<1000 Daltons) in a complex sample, such as biofluids (Beckonert et al 2007; Dona et al 2014). NMR-based metabolic profiling is a suitable tool for metabonomic applications as it provides a metabolic 'snapshot' representative of the sample at the time of the collection.

NMR spectroscopy is a powerful and versatile techniques which has been successfully applied in analytical chemistry, biochemistry, physiology and medicine. NMR has had huge progress in the understanding of the theory and the technical implementation over the last century. Therefore, NMR techniques can now be used to analyse liquids and solids to understand their chemical composition, molecular structure, and potential biochemical pathway.

2.1.1. FUNDAMENTAL PRINCIPLES OF NMR SPECTROSCOPY

Nuclear Magnetic Resonance (NMR) spectroscopy is based on the fundamental properties of the atomic nucleus which is composed of protons and neutrons. NMR detects quantum shifts using electromagnetic radiation interacting with the spin of atomic nuclei; this can be characterised by a nuclear spin quantum number, *I*. Some nuclei have integral spins (e.g. 1, 2, 3), others have fractional spins (e.g. 1/2, 3/2, 5/2) and some have no spin; all nuclei with a non-zero nuclear spin are able to be observed by NMR. Furthermore, many atoms spins are paired against each other, resulting in the nucleus of the atom having no nett spin. However, for biological purposes the nuclei with 1/2 spin

are the most useful and ^{1}H is best detected by NMR due to the high abundance of hydrogen nuclei in biological samples.

When a magnetic field is applied to these protons, their nuclei move from being in a random orientation to an ordered one circulating around the direction of the applied magnetic field. This circulation, or precession, causes a small magnetic field at the nucleus which opposes the externally applied field. The rate of the precession is known as the Larmour frequency of the nucleus and may be clockwise or anticlockwise but always the same for any given nuclei.

The external magnetic field causes the nuclei to separate into two different energy states, depending on their orientation and quantum spin number; in the case of the ^{1}H protons this would be high and low energy levels. The population of nuclei at each energy level can be calculated by the Boltzmann distribution.

$$\frac{N_\alpha}{N_\beta} = e^{\Delta E/RT}$$

Where N_α and N_β represent the number of nuclei in different energy states and R is the gas constant and T is the temperature. ΔE equals the difference or energy gap between the spin energy levels. This can be worked out using the equation below:

$$\Delta E = \gamma h B_0 / 2\pi$$

Where γ is the angular frequency (nuclear magnetic moment), *h* is the Plank's constant and B_0 is strength of the external magnetic field. Under the influence of the external magnetic field, nuclei at the lower energy level will leap to the higher enrgy level. The transition of nuclei back into the α energy level from the β state when the magnetic field is removed can be recorded and converted into an NMR signal. The difference between the energy levels is characterised by the resonance frequency. The energy difference represented by the intensity of the NMR signal is proportional to the total number of nuclei in the sample (Keeler, 2002; Ross et al 2007).

During an NMR experiment, a 90-degree radio frequency pulse is applied to excite the nuclei. The nuclei tip to the y axis and then when the radio frequency is turned off the nuclei will precess to the x axis; the preferred position (Figure 2-1). During this precession, an oscillating magnetisation voltage is released, called free induction decay (FID), and can be recorded. The FID represents the sum of the nuclei in the sample and can be transformed, using a Fourier transformation, to produce an NMR spectra with chemical shifts.

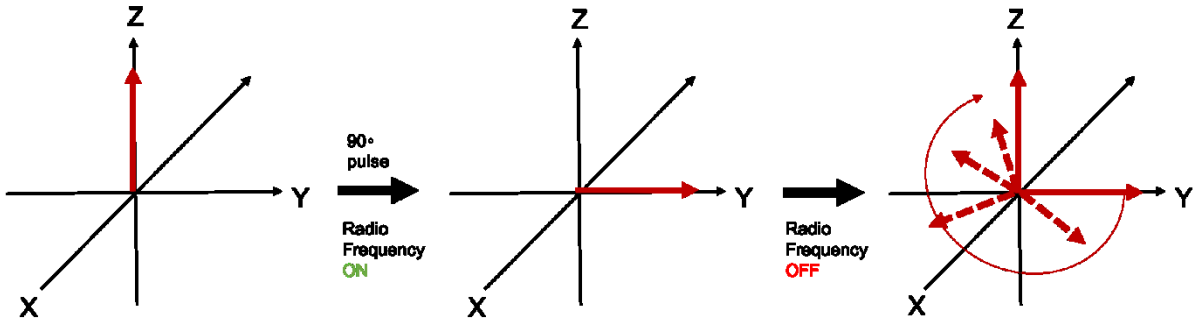


Figure 2-1. Representation of the 90-degree pulse effect on the nuclei. The nuclei at first are precessing around the external magnetic field (Z axis). When a 90-degree radio frequency pulse is applied and tips the nuclei to the XY plane instead. Once this radio frequency is turned off, the nuclei will precess back to the Z axis to reach equilibrium. This oscillating signal can be recorded and is known as the FID. (adapted from Keeler, 2002)

Nuclei experiencing the same chemical environment are called equivalent. The differences in chemical environment are represented on the NMR spectrum as chemical shifts on a parts per million (ppm) scale. Those nuclei experiencing different environment or having different chemical shifts are non-equivalent and nuclei which are close to one another exert an influence on each other's effective magnetic field. If the distance between non-equivalent nuclei is less than or equal to three bond lengths this effect can be observed; this effect is called spin-spin coupling. This coupling can cause the splitting of NMR peaks, and the shape of these peaks gives a further

indication of the number of neighbouring protons; this is called multiplicity, where peaks can be represented as singlets, doublets, triplets and multiplets.

One dimensional pulse sequences, with water pre-saturation, are typically used for metabonomic experiments on urine and faecal samples, as these samples are generally low in protein and lipid concentration (Beckonert et al., 2007). A standard ID experiment was used for all biofluids in this thesis and specific parameters can be found in the materials section of each results chapter; although other experiments including Carr-Purcell-Meiboom-Gill (CPMG) and diffusion edited experiments, which are typically used for plasma and tissue investigations.

2.1.2. TWO DIMENSIONAL NMR SPECTROSCOPY

Multi-dimensional NMR experiments are essential for additional structural information on molecules to help with identification, by plotting the data into two different domains. J-resolved (JRES) ^{1}H NMR spectroscopy allows the separation of chemical shifts and homonuclear J-coupling into different frequency domains. This is very useful to initially elucidate peak structure, for example whether two peaks are part of a doublet or two separate singlets.

Correlation spectroscopy (COSY) generates 2D NMR spectra in which signals of a normal single-pulse spectrum are correlated with each other; typically separated by two or three bonds. COSY experiments generate a spectrum that is presented by the chemical shifts from the nuclei as a diagonal line and the cross-peaks on either side of the diagonal that represent the coupling information between the nuclei.

2.1.3. LIQUID CHROMATOGRAPHY - MASS SPECTROMETRY

Mass Spectrometry (MS) is a very sensitive analytical technique which is typically coupled to upstream chromatographic platforms in metabonomic analyses. Unlike NMR, MS is a destructive technique since it requires the ionisation of the biological samples. Chromatography is used to

separate and fractionate the sample in order make the identification of metabolites easier. Metabonomic studies uses many different types of separation techniques, including liquid chromatography (LC), gas chromatography (GC) and capillary electrophoresis chromatography (CE). In this project, I will be focussing on LC-MS only, please refer to the literature for other separation techniques.

2.1.3.1. Principles of Mass Spectrometry

Mass spectrometry (MS) is an analytical tool in which molecules in ionised gaseous form are separated in space and time according to their mass to charge ratio (m/z) prior to detection. The MS is composed of the ionisation source, a mass analyser and an ion detector (Figure 2-2).

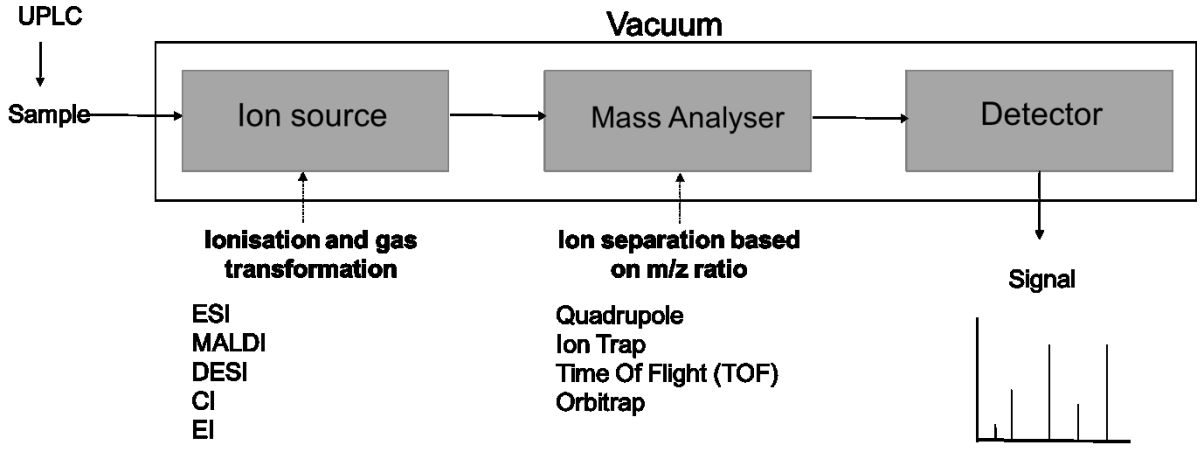


Figure 2-2: Diagram of mass spectrometer – the ionisation source, a mass analyser and an ion detector producing a signal. UPLC: ultra-pressure liquid chromatography, gas chromatography, ESI: electrospray ionisation, MALDI: matrix-assisted laser desorption ionisation, DESI: Desorption electrospray ionization, EI/CI: electronic or chemical ionisation

There are different ionisation techniques used for organic analysis including electrospray ionisation (ESI), matrix-assisted laser desorption ionisation (MALDI), desorption electrospray ionization (DESI), chemical ionisation (CI) and electro impact ionisation (EI) (De Hoffman et al 1996). For biological

investigations, ESI is commonly used as it is highly sensitive and it can directly convert ions from a liquid phase to a gas phase. ESI works by uses a capillary tip to produce a fine spray of charged particles, the solvent is evaporated by traveling through nitrogen gas and the charges enter the mass spectrometer (Pitt, 2009). ESI can produce positively or negatively ions and are analysed separately. By doing this, the number of molecules which can be identified can increase as some molecules will only be detectable in one polarity.

Once the molecules are ionised, they are transported through a vacuum into the mass analyser. This analyser separates the ions based on their mass to charge ratio (m/z). The mass analyser can be a quadrupole (Q), time of flight (TOF), ion trap or Orbitrap system. The quadrupole analyser uses four rods placed in parallel and a voltage is passed through them causing an electric field to be produced. This in turn allows the control of certain ions to pass through to the detector without hitting the quadrupoles (Pitt, 2009). The time of flight (TOF) analyser uses the fact that ions with a larger mass to charge ratio will travel through space slower than ions with a smaller mass to charge ratio to its advantage (Wiley et al 1955). The TOF will accelerate the ions through a high voltage causing further separation of ion species depending on their m/z ratio and helps with detection. The triple quadrupole (TQ) is typically used for MS/MS experiments to help with confident assignment of molecules using fragmentation. This fragmentation occurs in a collision cell placed between two quadrupole mass analysers (Pitt, 2009). Several analysers can be combined to approve ion separation and gain fragmentation information, including QTOF and the TQ. These are usually used in metabonomic studies to gain structural information through fragmentation patterns.

Finally, the ion species reach the detector, and are recorded by a current being produced when they hit the detector. The ion intensity and the m/z ratio is converted to make a digital signal and recorded by a computer (De Hoffman et al 1996).

2.1.3.2. Principles of Liquid Chromatography (LC)

Liquid chromatography (LC) is used to separate small molecules in a liquid mobile phase and passed through a column. These columns have different chemical compositions depending on the metabolite class of interest. Ultra-performance liquid chromatography (UPLC) uses smaller particles sizes (<2um) and operate at a higher pressure (>10,000 psi) to improve chromatographic performance compared to high performance liquid chromatography (HPLC); this includes peak resolution as well as increased sensitivity and speed (Wilson et al 2005).

A frequently used technique in metabonomic studies is reversed phase (RP) chromatography to separate medium to non-polar molecules. However, in this project, hydrophilic interaction liquid chromatography (HILIC) was used. This chromatographic technique is used for the retention of very polar molecules such as amino acids and organic acids (Spagou et al 2010). HILIC uses silica or a derivative of silica column and water in the mobile phase to further retain the polar molecules on the column (Dettmer et al 2007). The retention time vary between the substances so it can be used as an aid for the separation of molecules and therefore the identification of certain molecules.

2.2. MULTIVARIATE STATISTICAL ANALYSIS FOR METABONOMIC DATA

NMR and UPLC-MS data from biofluids are very complicated due to the complex nature of the mixture of chemicals and the heavy overlap of metabolite signals. Therefore, multivariate data analysis methods have been developed in order to extract significant information from the large data sets from these techniques. These techniques can simplify data sets and make visualisation of the similarities and differences between the different observations possible within biological samples.

An unsupervised multivariate technique commonly used is the principal component analysis (PCA), this establishes models with any *a priori* knowledge of class information in the dataset; thus,

indicating the intrinsic differences within the dataset. When focussing on group classification, supervised methods for data analysis can be used. These methods require prior knowledge of the class information and includes Partial Least Square (PLS) and Orthogonal Partial Least Squares - Discriminant Analysis (OPLS-DA). This class information means the data analysis optimises the separation between the different classes and requires subsequent validation via prediction of an independent test set to prevent over fitting of the data.

2.2.1. UNSUPERVISED METHODS

Principal components analysis (PCA) is a qualitative chemometric technique that reduces the number of variables that need to be considered into principal components that are linear combinations of the original variables (Wold et al 1987). This method is unsupervised, therefore no *a priori* knowledge of the sample classes is used and only the inherent variation in the data set is described. The data is firstly made into a 2D matrix of the different samples versus the number of variables or metabolic features.

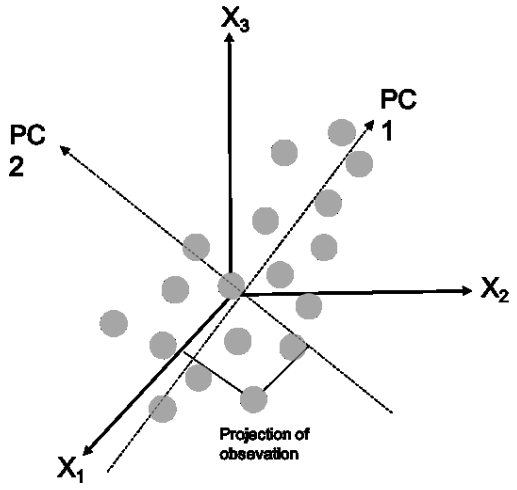


Figure 2-3: Representation of a PCA in three dimensions. The first two principal components (PC) describe the largest variance in the data. They produce a new plane for the data to be projected; presented as a scores plot.

By substituting the variables with derived axes called principal components, these latent axes correspond to the maximum variance of the dataset. The first principal component (PC) computed minimises the square of the distance of each data point to that vector. The first PC typically explains

the greatest variance in the matrix and the second principal component, which is orthogonal and uncorrelated to the first, will explain the second greatest variance (Wold et al 1987). Therefore, PCA allows the reduced multivariate data to be projected onto two or three dimensions.

This reduction in dimensions improves visualisation of the data as a whole and reveals patterns in the data; this visualisation is in the form of a scores and loadings plot. Scores illustrates the variability between the samples, whereas the loadings plot illustrates the weight of the variables causing this separation between samples. PCA is a useful analytical technique for metabonomic data as many metabolites are correlated so this information can be explained in less dimensions rather than looking at them individually.

2.2.2. SUPERVISED METHODS

Supervised statistical tools, for example projections to latent structures by means of Partial Least Squares (PLS) analysis and orthogonal partial least squares (OPLS), when the difference between two or more groups want to be investigated. The PLS method maximises the covariance between the metabolic data (X scores) and the thing we want to predict (Y scores); e.g. age, ethnicity or weight. This covariance is typically dominated by the X scores as this usually has a higher number of variables compared to the inner relationship of Y. Unlike PCA, this method maximises the covariance between the predictor of interest and response matrices rather than maximising the variance explained within the data. PLS can project the X- and Y-variables onto the same space within a scores plot to show the correlation between the co-ordinates of one observation on the X-plane to its corresponding co-ordinate on the Y-plane; this relationship is asymmetric (Tapp & Kemsley, 2009).

Orthogonal-PLS (OPLS) is an extension of PLS, which uses an orthogonal component to the model. Orthogonal variation means that the systematic variation in one block is not linearly related to the second block. This method can rotate the projection of a PLS plot to focus on the effect of interest

giving a better interpretation of the data. This is very useful for when you can't control for some variation or effects, which is typically useful for human or biological studies.

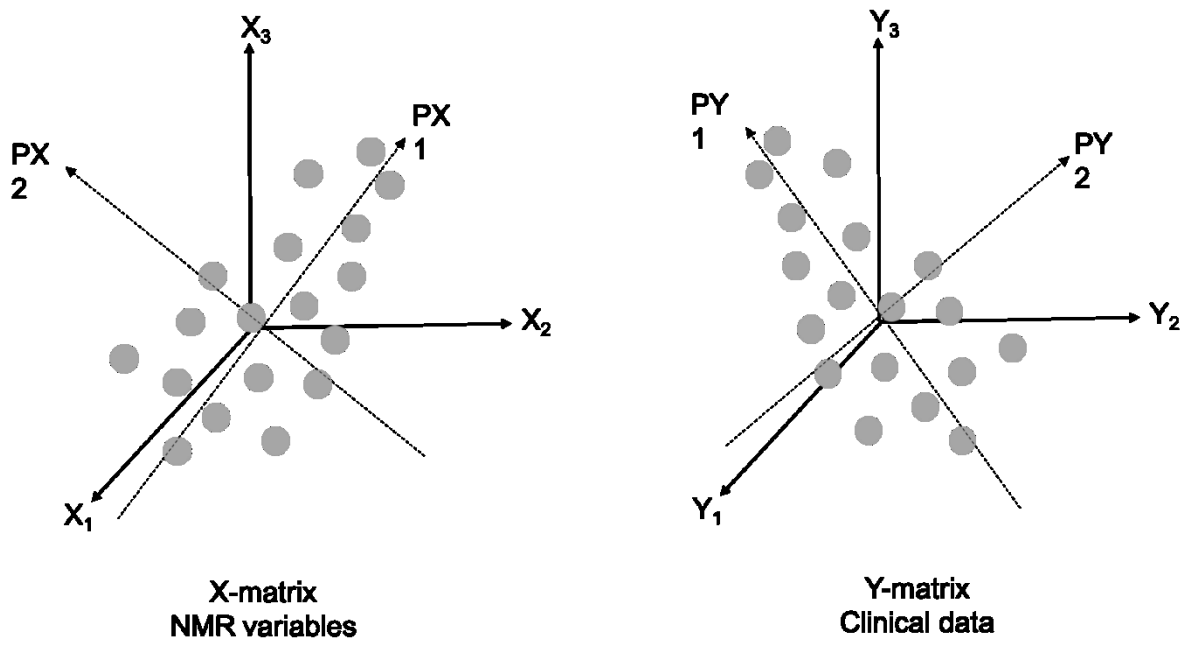


Figure 2-4 – Representation of PLS analysis. This is a regression extension of PCA and is used as a method to relate two data matrices together, in this case X = NMR variable and Y = clinical data (e.g. age or weight). PX and PY represent the scores matrices for the model.

OPLS-DA is an extension of OPLS and is used to obtain maximum separation between classes of data that are categorical and can be used to predict class membership. For OPLS-DA models, the X matrix is the sample NMR data whereas the Y matrix is made up of binary variables that describe the class for each sample. Even though it is similar to PCA, which describes the maximum variation within the data, it doesn't necessarily describe the maximum separation of classes in the PC scores plot.

A seven-fold cross-validation was also used to assess the robustness of the models and to ensure that the models weren't over-fitted. This test is done by removing a section ($1/7^{th}$) of data at a time and this is used to train the model and it then uses this information to predict the other sections of data. This is repeated for each section of data until all the data has been predicted. In order to

further test the validity of the model, permutation tests were performed. By assigning each sample with a random group and repeating the model 1000 times, a p-value can be produced by comparing Q^2Y values from the permutated models against the non-permutated test; a 95 % significance threshold ($p < 0.05$) was used for this project.

A heat map presenting the Pearson correlation coefficient values of the identified discriminatory resonances from the urine and faecal NMR and HILIC-MS data were constructed. This information is super imposed onto a pseudo NMR-spectra, or loadings correlation coefficient plot, which highlights the specific peaks causing the separation between the two classes (Cloarec et al 2005). The direction in which each peak, positive or negative, reveals the relative difference between the two groups.

2.3. METABOLITE IDENTIFICATION IN GLOBAL PROFILES

2.3.1. ^{1}H NMR DATA

Metabolite identification is achieved by first visualising specific data points in the NMR spectra. The designation of an identification was performed by comparing this information with available literature and existing databases, including in-house databases and the Human Metabolome Data Base (HMDB) found online. However, many molecule signals overlap in the NMR spectra making it hard to distinguish and make a conclusive assignment.

Therefore, statistical methods can be used to identify highly correlated data points in the NMR spectra to the variable in question and to extract structural information of the metabolite; this statistical method is called Statistical Total Correlation SpectroscopY (STOCSY) (Cloarec et al 2005). STOCSY is able to show highly correlated variables within in the whole NMR spectra and has the potential to highlight different molecules which are related to the same metabolic pathways. The correlation between NMR peaks to the selected variable of interest is indicated by a colour scale; ranging from low (blue) to high (red) correlation (Cloarec et al., 2005a). For this thesis, this method

was used as an aid to identify metabolites of interest within multivariate analysis as well as time series analysis.

2.3.2. HILIC-MS DATA

Using the correlation coefficient plots, the top 10 metabolites that caused the greatest separation in the groups was identified. The first step for identification was to search the chromatogram of the sample with the highest abundance of the molecule; using the median m/z value and retention time (RT) information. Once the exact m/z value, as well as information on adducts (if possible) was compiled from the sample. The designation of an identification was performed by comparing this information with available literature and existing databases, including METLIN from the Scripps Centre for Metabolomics and the Human Metabolome Data Base (HMDB) found online. MS/MS experiments were performed on QC samples only, therefore only molecules at a high abundance in these samples had fragmentation information available to aid identification of features.

The level of identification was determined using the criteria found in Sumner et al (2007); level 4, an unknown compound that couldn't be identified; level 3, a known compound class through the identification through database searching ; level 2, the mass to charge ratio for a particular molecule matches the databases; level 2+, the mass to charge ratio and MS/MS fragmentation data is matched to the databases; level 1, the compound has been identified and confirmed using chemical standards (Sumner et al 2007).

2.4. TIME SERIES ANALYSIS FOR ^{1}H NMR SPECTRAL DATA

In this project, a time series analysis has been used to identify metabolites, which are different between groups through time using an in-house time series algorithm (Wolfer et al., In Press). This method of analysis was used to further understand specific metabolite trajectories in infant faecal

samples in the first two years of life. This technique relies on functional spline for short asynchronous time series analysis and functional data analysis (FDA) (Ramsay et al 1988).

A variable of interest should differentiate groups through time (e.g. term vs. preterm born). This corresponds to a variable for which the time-trajectory of term samples if significantly different from the preterm samples. To evaluate the significance of the actual difference between these group trajectories, these group curves need to be compared (and out-perform) to randomly generated groups (i.e. term or preterm).

For this method, ^{1}H NMR data was binned into 1000 bins; variables represent 5 columns of NMR data summed and assigned a ppm region at a distinction of 0.01ppm. For each individual, each variable (binned ^{1}H NMR data) is plotted onto a graph and a curve is used to best represent the data trajectory through the time points. The benefit of using curves to compare between different individuals rather than specific data points is that there can be missing values from individuals without affecting the curve. The smoothness of this curve, using 4 degrees of freedom, corrects for the measurement of error but will not over fit the data (Figure 2-5(a)). Therefore, a curved line that doesn't go through all data points is used to describe the average rather than a straight line (similar to a linear regression).

For the different groups, an average or mean group curve is produced; in this project, this included term *vs.* preterm at birth, mode of delivery and BMI status at 1 and 2 years of age (low versus high). Group curves have 95% confidence bands, constructed from a bootstrapping technique; this technique randomly assigns individuals in a group and reruns the algorithm another 1000 times. An empirical p-value can be determined by comparing the actual value with the null distribution of the value which has been permuted.

Furthermore, to evaluate the significance of the difference between the two mean group curves (e.g. term vs. preterm born) a permutation test is performed. This test assigns each individual with a

random group and reruns the algorithm 1000 times to calculate a p-value for the specific variable. The p-value is calculated by how many times the random assignment of groups is superior to the actual assignment of groups and then divided by the number of tests performed; in this case 1000 times.

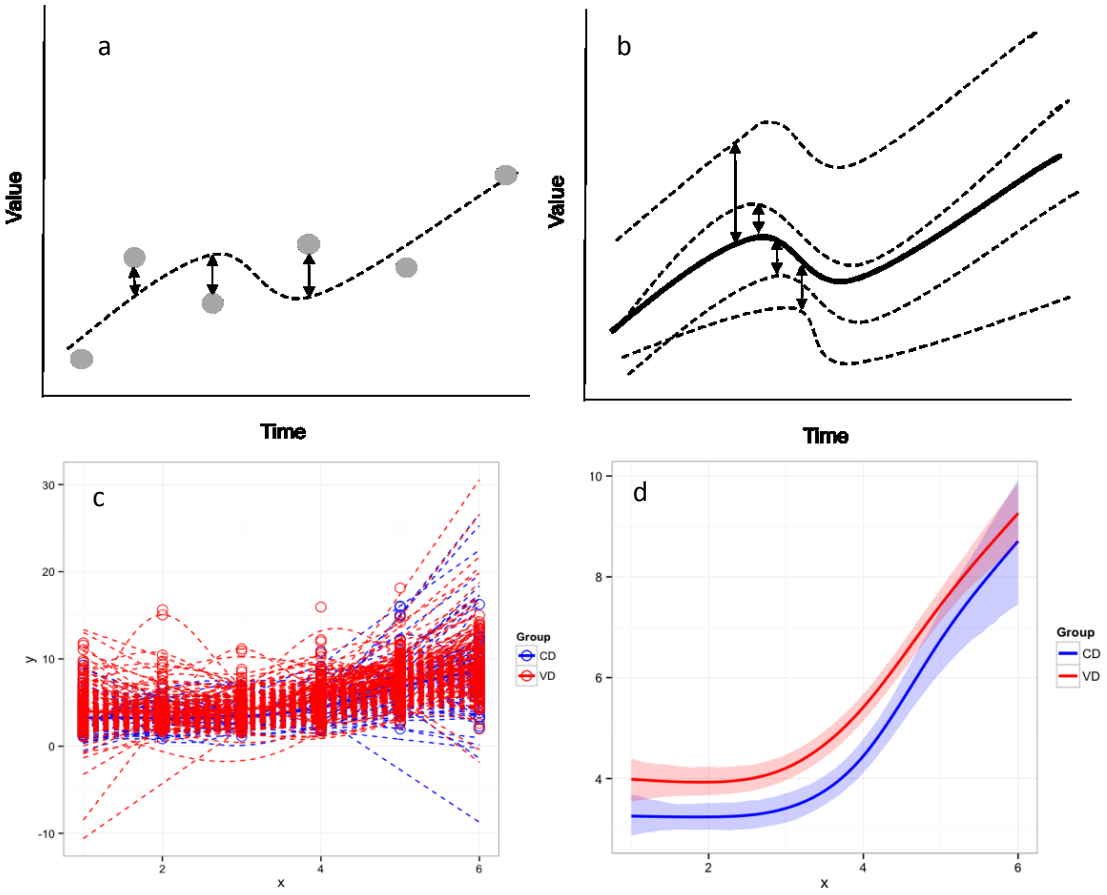


Figure 2-5: Representation of time series analysis. a) the curve (dotted line) is fitted for each variable for each individual, with arrows representing the measurement of error. b) curves from each individual for each variable is compared and compiled with individual variability accounted for (arrows). c) groups are represented by different colours, individual measurements by circles, individual trajectories by dashed lines, d) group mean curves by continuous lines and 95% pointwise confidence area on the group mean curve estimation is drawn as a coloured area.

As the p-value is obtained or each variable (ppm) a false discovery rate (FDR) correction can be applied across the resulting p-values. The algorithm applies a Benjamini-Holberg FDR correction to

the obtained p-values. As a p-value of zero should not be reported, the lowest p-value obtainable (with the present settings) when no random groups out-perform the actual groups is 0.000999.

By examining the spline graphs, the bold lines represent the group mean curves and the 95% pointwise confidence area on the group mean curve is the shaded area in the same colour (Figure 2-5(d)). This graph and the corresponding p-value informs the difference in group's specific variables or metabolites at certain time points. Once these variables were identified, STOCSY (section 3.1) was used on the NMR spectral data to assign and identify specific variables to metabolites.

2.5. GUT MICROBIAL PROFILING

Microbiome-host interaction has been found to play a crucial role in mammalian health and disease development. Therefore, studying the microbial communities present in various locations of the human body has contributed to our knowledge of this interaction. Metagenomics is defined as the culture-independent study of sets of genomes from a microbial population or the exploration of the entire microbial genome; this field has expanded exponentially due to the growing interest in of the role of the microbiome in human health. As over 80% of the bacteria in the gut are uncultivable it means the improved DNA-sequencing technologies have enabled enhanced capabilities for sequencing large scale data sets. However, the complexity of the microbial communities makes this field very challenging. The genetic information of the metagenome can be up to 100 times greater than the human genome.

2.5.1. 16S RIBOSOMAL RNA GENE SEQUENCING USING ILLUMINA MISEQ SEQUENCING

The 16S ribosomal RNA (rRNA) gene codes the 16S rRNA, which is a part of small subunit (30S) of the prokaryotic ribosome (Woo et al., 2008). The ribosome is an organelle that performs translation or protein synthesis; it consists of a large and a small subunit. The small subunit binds to mRNA and the

large subunit catalyses protein bond formation between amino acids to produce peptides. The 16S rRNA gene codes for the small subunit of a prokaryotic ribosome; this is homologous to the eukaryotic 18S rRNA. Therefore, this gene is useful for sequencing in different types of samples, especially human biofluids, as it is unique to prokaryotes.

The 16S rRNA gene is very well conserved between different species due to its important function in the cell so it is an ideal gene for sequencing to differentiate between bacterial species. However, between the conserved regions, there are 9 variable (V) regions within the gene, which are different between species of bacteria; therefore, specific bacterial species can be identified. This technique was used for the creation of the first phylogenetic to understand evolutionary relationships between species (Fox et al 1977). Using this gene, different types of sequencing technologies have been produced including Sanger's sequencing (Sanger, Nicklen, & Coulson, 1977), 454 Roche sequencing (pyrosequencing) (Margulies et al., 2005), and MiSeq (Illumina). For the purpose of this thesis, I will be discussing the MiSeq technology only, please refer to the literature for more information on other technologies.

MiSeq by Illumina applies amplicon sequencing to analyse genetic material; in this case the 16S rRNA gene sequence. The forward and reverse universal primers, which target the conserved region of the gene, are used to make a template from the DNA strand for both strands separately. A single-strand of DNA is attached to the surface of a matrix using an adapter at both ends of the DNA fragment. Fluorescently-labelled nucleotides are added to the DNA strand one at a time, and as they attach to the daughter DNA strand a fluorescent signal is given off. The fluorescent signal omitted for each nucleotide is different and distinguishable so the signals can be determined by a fluorometer and recorded as a digital output (Clarridge, 2004; Mardis 2008).

2.5.2. PRE-PROCESSING 16S rRNA SEQUENCED DATA

DNA was isolated from faecal pellets using PowerFaecal DNA Isolation kits (MoBio Laboratories Inc) in accordance to the manufacturer's protocol and eluted in the 50 μl of water. This was sent for sequencing using the V4 region of the bacterial 16S rRNA gene using Illumina by Research and Technology (USA).

Using this method of sequencing produces two reads for each strand of DNA (forward and reverse) and needed to be put together. This was done using the Mothur programme and standard scripts associated to this programme (Schloss et al 2009). This data was aligned and clustered together depending on how similar to each other they are. The sequences are then classified according to the 16S rRNA gene and only bacterial species data is kept; eurkaryotic and archaea data is removed. The sequences are then clustered into Operational Taxonomic Units (OTU) and the taxonomy for each OTU was matched. The data output is a list of bacteria species with Phylum, Class, Order, Family and Genus information for each OTU as well as the size or number of times each OTU was in a sample. In this project, 569 OTUs or different species of bacteria were recognised from the faecal samples. When processing, 12 samples which had less than 2000 sequences were excluded from the analysis. This was done because samples with low number of sequences would reduce the percentile of OTUs and therefore reduce how detailed the outcome could be.

2.5.3. ANALYSIS OF MICROBIAL PROFILES

The data, in the form of OTUs, taken from Mothur was then imported into the statistical analysis of taxonomic and functional profiles (STAMP) software, with the corresponding metadata for each sample, to analyse the microbial profiles (Parks et al 2014). This software was used to identify features that differ significantly between microbial profiles from different groups; in this project the difference between infants born term and preterm. As the data was not normally distributed, non-parametric tests were used. For the two groups, the two sided White's parametric t-test was

performed on all identified bacterial species in the sample. In order to avoid false positive results in the data analysis, the Benjamini-Hochberg correction was applied using this software. STAMP software produced extended error plots to show significantly different bacterial phyla, classes and genera in samples taken from infants born term or preterm at different time points. These extended error plots also include 95% confidence intervals. Only OTU differences between samples with a p-value of less than 0.05 was used as a cut off for significance.

CHAPTER 3

METHOD DEVELOPMENT OF INFANT BIOFLUID SAMPLE PREPARATION FOR METABOLIC PHENOTYPING

3.1. INTRODUCTION

Plasma, urine and faecal water metabolic fingerprints capture essential information associated with host energy metabolism and microbial-host co-metabolism (Nicholson et al 2012). The majority of biofluid sample preparation methods for metabolic phenotyping have been produced for adult studies, but pipelines created especially for new-born babies and infants are lacking (Jacobs et al 2008; Monleon et al 2008; Pettersson et al 2008; Saude et al 2007). Infant biofluid samples are often challenging for metabolic phenotyping, mainly due to their smaller volumes compared to adults; 24 hour urine volume for new borns is 20-350mL in comparison to a normal adult, where the urine volume ranges from 750-2000mL. This is especially true for blood samples, where infants undergo a heel prick and blood is collected onto filter paper (Pitt 2010). Whilst in hospital, urine samples can be collected using plastic collection bags attached with adhesive, but typically cotton balls are placed inside disposable nappies for collection. Faecal samples are collected in a similar manner to urine samples, through disposable nappies, which causes concern for cross contamination of urine and faeces. In early life, depending on the diet of the infant, faecal samples can be liquid in consistency, adding to the challenges of biofluid collection. Furthermore,

contamination introduced from disposable nappies and cotton wool is a major consideration, which may introduce biases in downstream 'omic methods (Goodpaster et al 2011).

It is thus clear that there is a critical need to optimise sample preparation and data acquisition protocols for the infant clinical samples. Firstly, the effect of dilution factor on biosamples (including, urine, plasma and faecal water; dilution series concentrations: 0, ½, ¼, ⅛) for metabolic phenotyping, using ^{1}H NMR spectroscopy was investigated. Following this analysis a method for faecal water extraction was established and optimised, using a range of different extraction techniques and solvents, for metabolic phenotyping using ^{1}H NMR spectroscopy. Finally, a database of contaminants associated with infant hygiene products including nappies and cotton wool was established using ^{1}H NMR and UPLC-MS.

3.2. AIM

The aim of this work was to evaluate and optimise infant biofluid (urine, plasma and faecal) sample preparation methods for metabolic profiling using ^{1}H NMR spectroscopy and UPLC-MS, combined with multivariate data analyses. I will be fully evaluating any contaminants coming from cotton wool and nappies during the collection of urine samples. Furthermore, due to the large amounts of faecal samples being used in this project, the optimisation of faecal water preparation in infants was done.

3.3. MATERIALS AND METHODS

3.3.1. SAMPLE DONOR INFORMATION AND ETHICS

Urine and faecal samples were collected from five term born infants, further information can be found in Table 3-1. A plasma sample was collected from five healthy adult donors (due to ethical considerations with respect to sampling of neonates). Ethical approval for the use of these samples was granted from the West London Research Ethics committee (REC reference number 12/LO/0203 and 10/H0/7135).

Table 3-1. Clinical information on infant urine and faecal samples used in method development for dilution series analysis. (T: Term, P: Preterm, M: male, F: female, VD: vaginal delivery, CS: caesarean section)

Clinical data	Urine					Stool				
	1	2	3	4	5	1	2	3	4	5
Term or Preterm	T	T	T	T	T	T	T	T	T	T
Gestation age at birth (weeks)	40.4	41.4	41.8	40.7	41.7	40.1	37.1	39.1	39.3	40.4
Weight (kg)	3.75	3.34	3.93	3.96	3.6	3.69	2.8	3.24	2.7	4.07
Sex	F	M	M	M	M	M	F	M	M	F
Mode of delivery	CS	VD	CS	VD	VD	VD	VD	CS	CS	VD
Sample week PP	1	1	0	0	1	0	1	0	2	1

3.3.2. DILUTION FACTOR OF URINE, PLASMA AND FAECAL SAMPLES

Urine and faecal samples were collected from five term born infants and a plasma sample was collected from five healthy adult donors (due to ethical considerations with respect to sampling of neonates). A dilution series was produced for each biofluid prior to profiling by ^{1}H NMR spectroscopy. For each of the samples, a dilution series (0, ½, ¼, ⅛) was created (see supplementary data (S1) for further information). To compensate for the decrease in signal to noise in more dilute samples, the length of the acquisition was increased, acquisition scans quadrupled as the dilution

increased, which in turn increased the duration of the NMR experiment. Samples were also subject to the same number of NMR acquisitions (32 scans).

3.3.2.1. Urine Preparation for Metabolic Profiling

For the undiluted samples, 540µl of urine was mixed with 60µl phosphate buffer solution at pH 7.4 (0.6M Potassium Dihydrogen Phosphate (KH_2PO_4), 0.01% 3-Trimethylsilyl Propionic Acid-d4 Acid Sodium Salt (TSP), 3mM Sodium Azide (NaN_3), and Potassium Hydroxide Solution KOH 45% w/w in 80% D_2O). After centrifugation at 16000 *xg* for 10 minutes at 4°C, samples were transferred to 5mm outer diameter NMR tubes. Depending on the dilution series, samples were mixed with water to make up the volume up to 600µl, the same amount of phosphate buffered solution with the internal standard, TSP, was added to each sample.

3.3.2.2. Plasma Preparation for Metabolic Profiling

For the undiluted sample, 300µl was mixed with 300µl of phosphate buffer solution. After centrifugation at 16000 *xg* for 10 minutes at 4°C, samples were transferred to 5mm diameter NMR tubes. Depending on the dilution series, samples were mixed with water to make up the volume up to 600µl; the same amount of phosphate buffer solution was added to each sample.

3.3.2.3. Faecal Water Preparation for Metabolic Profiling

From the native sample, 200mg of faecal sample was homogenised (*vide infra* section 3.4.2) with 500µl of water using zirconium beads. The samples were centrifuged at 16000 *xg* for 10 minutes at 4°C and faecal water (supernatant) was obtained. This suspension was mixed with 60µl of phosphate buffer solution and samples were transferred to 5mm outer diameter NMR tubes. To make this dilution series, different masses of faecal matter was mixed with the same amount of water (500µl) and the same amount of phosphate buffered solution was added to each sample.

3.3.3. ASSESSMENT OF AGE-RELATED VARIABILITY ON INFANT URINE PROFILES

Infants develop rapidly from birth, due to interactions with their internal and external environments and as their gut microbiome starts to develop and this is likely to cause their metabolic profile to dramatically change in the first few weeks of life. Therefore, the assessment of age variability on infant profiles was investigated for metabolic phenotyping by ^{1}H NMR spectroscopy. Urine samples taken for mother-infant comparison were obtained from a mother-infant matched samples using an infant born at term. Urine samples were taken from the infant at 0, 14 days, 8 weeks and 14 months' postpartum and a further maternal sample was taken at the time of birth for comparison**.**

3.3.4. SAMPLE COLLECTION MATERIAL CONTAMINATION ASSESSMENT

To assess contamination of infant urine metabolic profile introduced by collection protocols, different nappies and cotton wool brands were investigated. Three brands of nappies were assessed; Pampers, a UK hospital brand (Libero) and a UK store brand (Tesco). Four cotton wool brands were also assessed: three UK store brands (Tesco, Sainsbury's and Boots) and a UK hospital brand (Robinson's healthcare). Chemical profiles arising directly from these products were assessed using H_2O as a blank sample, as well as a pooled infant urine sample from infants at postpartum ages from 0-19 weeks.

A volume of 3 mL of urine was added to cotton wool and incubated for two hours at 37°C (to mimic the typical infant urine collection conditions); the cotton wool was either frozen overnight at -80°C prior to extracting urine or urine was extracted directly from the cotton wool immediately after collection.

Extraction from the cotton wool was carried out in two ways to assess whether chemical contamination was from exposure to laboratory equipment: 1) squeezed by hand from the cotton wool into a falcon tube or 2) extracted by centrifugation at 16000*xg* for 5 minutes. Extraction by centrifugation involved adding the cotton wool to a falcon tube and positioning a microcentrifuge

tube underneath it to stop the cotton wool from moving to the bottom of the tube to ensure as much urine was extracted as possible (Supplementary – Figure S2).

It was not possible to gain urine from nappies directly due to the absorbent core of the nappy. Therefore, nappies were also assessed using a similar method to the cotton wool; urine was added to cotton wool and this was inserted into the different nappies and incubated at 37°C for two hours. Urine was extracted from the cotton wool by manual squeezing.

For ^{1}H NMR analysis, 540µl of urine and 60µl of phosphate buffer was transferred to 5mm NMR tubes and acquired. The remaining urine was added to two well plates and run by both Reversed Phase (RP) and Hydrophilic Interaction Chromatography Mass Spectrometry (HILIC) in both electrospray ionisation positive and negative mode. In addition to urine, water was used to obtain an experimental blank sample from both cotton wool and nappies. These were treated in the same way as urine samples for preparation and analysis.

3.3.5. FAECAL WATER EXTRACTION OPTIMISATION PROCEDURE

The schematic representation for optimal faecal water extraction experimental design during this method development can be seen in Figure 3-1 below.

3.3.5.1. Freeze-thaw cycle effects

It has been previously postulated that by subjecting a sample to freeze thaw cycles, it will increase the concentration of metabolites released into the water (Wu et al 2010). Therefore, the influence of freeze thaw cycles was assessed using the same faecal sample with 200mg aliquots taken after being subject to different conditions. These conditions included: no thawing of the sample, 1, 2, and 3 freeze-thaw cycles; for each freeze cycle, the samples were subject to freezing at -80°C for two hours. All aliquots were added to 600µl of water, vortexed for 15 seconds and sonicated for 30 minutes. Samples were centrifuged for 20 mins at 13,000 *xg* and 540µl of supernatant was mixed

with 60µl of phosphate buffer and transferred into a 5mm NMR tube. ^{1}H NMR spectra were acquired for all the aliquots via the method described in section 3.5 of this chapter.

3.3.5.2. Homogenisation techniques

Due to the nature of faecal samples, it is important to fully homogenise the sample to ensure that all metabolites are released into the faecal water. Four homogenisation techniques were assessed: 1) bead beating, 2) vortexing, 3) homogenisation by pestle, and 4) homogenisation by liquid nitrogen. The experiment was conducted in duplicate and faecal samples were, firstly, weighed out (200mg).

The samples that were homogenised with beads and vortexed, had 500µL of water added prior to homogenisation, whereas pestle and liquid nitrogen homogenisation samples were mixed with water and vortexed for 1 minute after the homogenisation technique. 1) For the bead beating technique, faecal samples were combined with 0.1mm of 1mm Zirconium beads and bead beat for 30s using the Biospec Mini Beadbeater 96, 2) for the vortexing technique, samples were vortexed for 1 min and sonicated for 5 minutes, 3) samples underwent homogenisation in microcentrifuge tubes using a plastic pestle and 4) samples were placed in a mortar and liquid nitrogen was poured into the mortar to deep freeze, and the a pestle was used to produce a dehydrated powder. All samples were centrifuged for 20 minutes at 13000 *x g* to obtain a supernatant of faecal water.

Further to this, a raw sample that was not homogenised and only centrifuged to obtain as much natural faecal water as possible, and named the centrifuge only technique. This faecal water (400µL) was mixed with phosphate buffer (200µL) and transferred to 5mm NMR tubes; ^{1}H NMR spectra were acquired using the experimental parameters described in section 3.5.

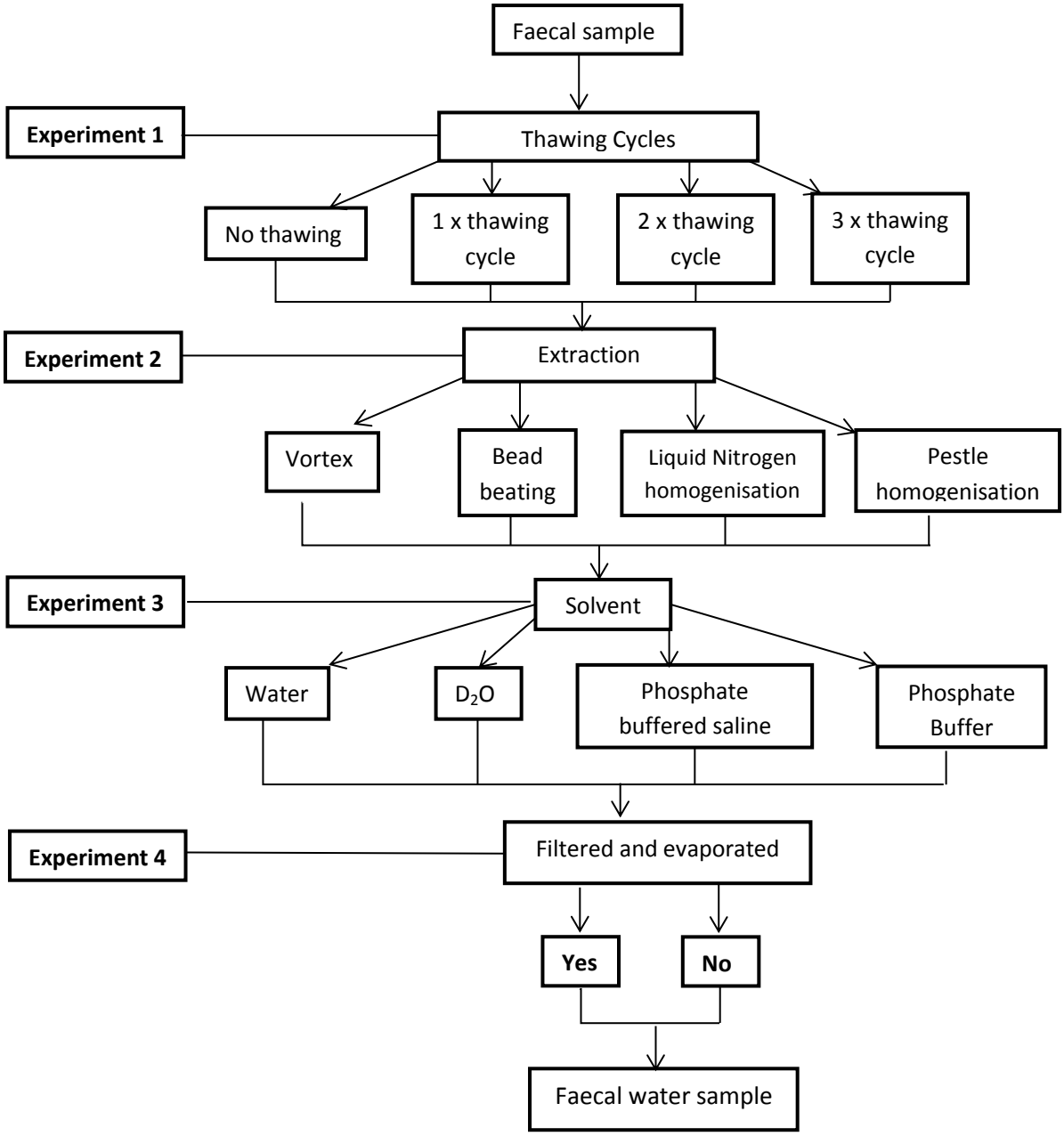


Figure 3-1: Schematic representation of the summary of experimental design for infant faecal water extraction optimisation.

3.3.5.3. Solvent extraction

Use of the appropriate solvent is important for the full retrieval of information from the stool sample and to optimise the resolution of spectroscopic analysis (Jacobs et al 2008). Four different solvents were assessed: 1) deionised H_2O, 2) heavy water - D_2O, 3) PBS (1.9 mM Na_2HPO_4 , 8.1 mM NaH_2PO_4 , 150 mM NaCl, pH 7.4) and 4) Phosphate buffer (KH_2PO_4, TSP, NaN_3, and KOH 45% w/w, in D_2O).

Faecal water was extracted in duplicate from each sample to assess the influence of different solvents on the metabolic profile of faecal water by ^{1}H NMR spectroscopy. Three faecal samples were divided into four 200mg fractions each. Each fraction was mixed with 400µL of different solvent, homogenised by vortexing and sonication and centrifuged for 20mins at 13.000 × *g*. From the supernatant, 200µl of faecal water was mixed with 340µl of D2O and 60µl of phosphate buffer; this was transferred to a 5mm NMR tube; ^{1}H NMR spectra were acquired using the experimental parameters described in section 3.5.

3.3.5.4. Optimisation for UPLC-MS analysis

Due to the small mass that is typically available from infants, an optimised technique for the preparation of faecal samples for both ^{1}H NMR and UPLC-MS acquisition was needed for this project. Therefore, for UPLC-MS, an organic solvent is needed to precipitate proteins from the sample and improve the retrieval of metabolites, especially species of bile acids (Michopoulos et al 2009; Sarafian et al 2015).

Furthermore, UPLC-MS proposes to filter samples through 0.2 µm Costar's Spin-x centrifuge tube filters (Sigma CLS8160) to ensure larger particles in the sample are removed and don't block or damage the column.

Two aliquots from the same faecal sample (200mg) was added to acetonitrile (ACN) and UPLC-grade water (500µl) at a ratio of 1:3 (ACN:H_2O). Another two aliquots of the same faecal sample was also extracted using UPLC- grade H_2O only was compared with samples extracted using the ACN:H^2O mix. The samples were then homogenised using the bead beating technique and centrifuged following the protocol described in section 3.4.2. of this chapter.

To compare filtering of samples, one sample from both solvent types had the supernatant transferred into the filters and centrifuged for a further 20 mins at 16000 *xg*. The other two faecal samples from the different solvent types, weren't transferred into filters and just centrifuged for the same amount of time. Faecal water that had successfully passed through the filter was then aliquoted into two separate vials; for ^{1}H NMR analysis and UPLC-MS analysis. As these samples were being acquired using ^{1}H NMR analysis, the ACN was evaporated using speed vacuuming for 30 mins and re-suspended in 600µl of phosphate buffer. This was transferred to 5mm NMR tubes and ^{1}H NMR spectra were acquired using the experimental parameters described in section 3.5.

3.3.6. ^{1}H NMR SPECTROSCOPY AND DATA PRE-PROCESSING

^{1}H NMR spectra were acquired on a Bruker DRX-600 spectrometer (Bruker Biospin, Karlsruhe, Germany) operating at 600.29MHz for proton observation using a standard one dimensional water pre-saturation pulse sequence [relaxation delay-90°-t1-90°-tm-90°-acquire free induction decay (FID)]. The relaxation delay was 4 seconds, with application of a 90° radio frequency pulse, t1, referring to the interpulse delay, which was set to 3µl, while tm is the mixing time of 100ms. The probe was matched and tuned automatically to the proton transmitter resonance frequency before acquisition for each sample and samples were run at a temperature of 27°C (300K).

Processing of ^{1}H NMR spectra was carried out using TOPSPIN 3.1 software package (Bruker Biospin, Rheinstetten, Germany). The FIDs were transformed into a spectrum by Fourier transformation. The

spectra were manually phased, baseline corrected and calibrated to the TSP signal at δ0.0 for the urine and faecal samples and the glucose signal at δ5.25 for plasma samples.

3.3.7. UPLC-MS GLOBAL PROFILING

3.3.7.1. RP-UPLC-MS Analysis

UPLC separation was conducted using an Acquity UPLC System (Waters Corp, USA) using an Acquity UPLC CSH C18 2.1 × 100 mm, 1.7 µm, column (Waters Corp, USA). Column temperature was set at 45 °C and flow rate at 0.5 mL/min. Mobile phase A consisted of 0.1% formic acid in water and mobile phase B of 0.1% formic acid in ACN. The elution gradient was set as follows: 99% A (0.0−3.0 min), 85% A (3.0−6.0 min), 50% A (6.0−9.0 min), 5% A (9.0−10.1 min; curve 1), 99% A (10.1−12 min). Injection volumes of 1 µL were used for both positive and negative ionization modes. The autosampler was set to 4 °C. Mass spectrometry was performed using a Xevo G2 QToF (Waters MS Technologies, U.K.) with an electrospray ionization (ESI) source. A Quality Control (QC) strategy was employed for the UPLC-MS analysis. This QC was produced from a pooled sample of all the urine extracts collected from the cotton wool balls and nappies. This sample was injected at least 10 times before initiating the run, in order to condition the column. The QC was subsequently reinjected at the beginning of the run, every 5 sample injections, and finally at the end of the run to assess instrument stability and analyte reproducibility. H_2O experimental blanks as well as a pooled infant urine sample were injected at the very end of the run.

3.3.7.2. HILIC-UPLC-MS Analysis

Acquity UPLC BEH HILIC 2.1 × 100 mm, 1.8 µm, column (Waters Corp, USA) was also performed to maximise global profile coverage of metabolites. Column temperature was set at 40°C. Mobile phase A consisted of acetonitrile (ACN)/water (95:5) and mobile phase B ACN/water (50:50). In both solutions ammonium acetate was diluted to 10 mM final concentration (5%). The elution gradient

was set as follows: 99% A (0.0−2.0 min; 0.4 mL/min), 99−45% A (2.0−8.0 min; 0.4 mL/min), 45−1% A (8.0−9.0 min; 0.4 mL/min), 1% A (9.0−9.1 min; 0.4−0.6 mL/min), 1% A (9.1−11.1 min; 0.6 mL/min), 1−99% A (11.1−11.2 min; 0.6 mL/min), 99% A (11.2−17.6 min; 0.65-0.9 mL/min), 99% A (17.6−17.7 min; 0.7−0.5 mL/min), 99% A (17.7−20.0 min; 0.4 mL/min). An injection volume of 1 µL was used for both positive and negative ionization polarity modes. The autosampler was set to operate at 4°C. Mass spectrometry was performed using a Xevo G2 Q-TOF (Waters MS Technologies Ltd., UK) with an electrospray ionization (ESI) source. Finally, the same QC strategy was followed as for the RP-UPLC-MS analysis with H_2O experimental blanks and a pooled infant urine sample being injected at the end of the run.

3.3.8. DATA PROCESSING AND MULTIVARIATE DATA ANALYSIS

3.3.8.1. ^{1}H NMR Data processing

The spectral data (from δ0.5 to 9.5) were imported into Matlab software (version 2014a, the Mathworks Inc, MA, USA) and were transformed into 32K data points. Resonance of the water (δ4.7 -5.05) was also removed from each spectra. Each ^{1}H NMR spectrum was aligned using an in-house algorithm and normalised using Probalistic Quotient Normalisation (PQN) in order to remove variation in metabolite linked to osmolarity and dilution (Dieterle et al 2011).

3.3.8.2. Data Extraction for UPLC-MS

Data collected following HILIC and RP-UPLC-MS analysis (both positive and negative) were subjected to peak-picking and grouping using MarkerLynx XS (Waters Inc., v4.1) software. Samples were normalized to median fold change and filtered once again using coefficient of variation in the QC samples; this removed features with low repeatability within the QC samples run throughout the experiment. The remaining features were produced into a table and imported into SIMCA-P 14 software (Umetrics, Umea, Sweden). The most relevant features or metabolites in pair-wise models

from MS data were flagged by using S plots; the top 10 significant features which cause the separation in the model was taken to be identified.

3.3.8.3. ^{1}H NMR and UPLC-MS metabolic data analysis

Multivariate statistical models (PCA, OPLS, OPLS-DA) were performed in Matlab software (version R2014b, the mathworks Inc, MA USA) and SIMCA-P 14 software (Umetrics, Umea, Sweden) using UV scaling. The Q^2Y value was determined by plotting the score against the cross validating scores to estimate the goodness of fit and level of over fitting of the OPLS and OPLS DA models. All models were validated by multiple permutation tests (permutation number =100). (see Methods Chapter for further details)

3.4. RESULTS

3.4.1. ^{1}H NMR SPECTROSCOPY BIOFLUID METABOLIC PROFILES WITH DILUTION FACTORS

3.4.1.1. Urine Dilution Series

The urine ^{1}H NMR spectra obtained from an individual infant, with each spectrum corresponding to a different dilution factor and run at the same number of acquisitions are shown in Figure 2a. Samples were both subject to the same number of NMR acquisitions as well as increasing the number of acquisitions to correspond with the dilution series. To adjust for this difference in acquisitions and therefore intensity of signals, the same amount of the internal standard, TSP, was added to all of the samples. By scaling the NMR spectra - which had increasing scan length- to the integral of the TSP peak at δ0.0, this correctly adjusted and calibrated the spectra of all of the samples to its relative intensity (supplementary - Figure S3) (Spagou et al 2010; Hoult et al 1976).

These results indicate that after adjustment for the number of scans to account for differences in signal to noise, there was little visual effect of dilution on the NMR spectral signal intensities (Figure 3-2b). The main difference seen in samples is an increased peak at δ1.25, which has been identified as the contaminant tert-butanol. This molecule has been seen in a number of different cases within the department including from the nappy contamination, which is discussed later in section 4.3.1 of this chapter. It is believed that this molecule is present in the phosphate buffer, which is used to balance the pH in biofluids for NMR analysis. However, as all of the samples had the same amount of phosphate buffer added before analysis but variable amounts of distilled water, it can be concluded this contaminant is present in the water also.

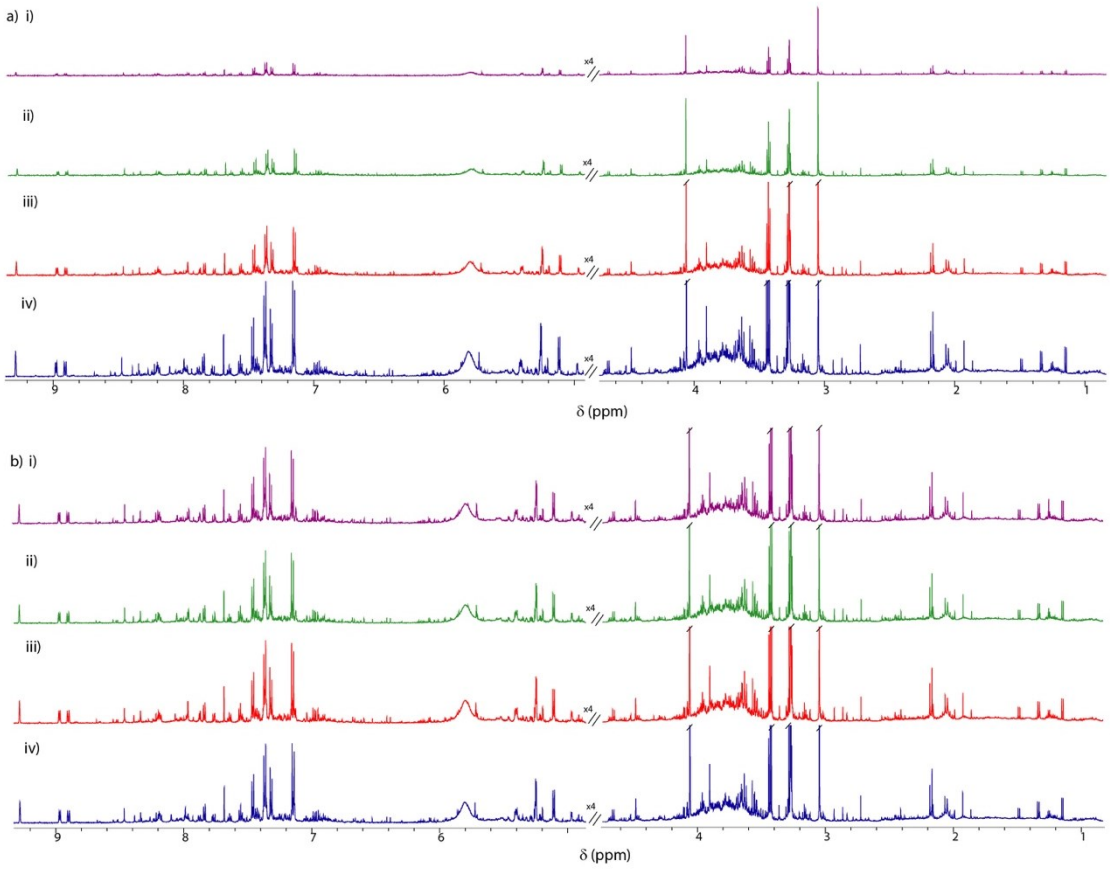


Figure 3-2 a) 1D ^{1}H NMR urinary dilution series spectra from one infant, with the same number of acquisitions (32 scans). b) 1D ^{1}H NMR spectra from one infant with dilution series with increasing number of acquisitions as dilution increases. Key: Dilution series: i)68µl, ii) 135µl, iii)270 µl, iv) 520 µl of urine

The dilution factor was shown to affect the computational models derived from the spectral data when non-normalised spectra were used, suggesting that dilution has a physicochemical effect on the observation of certain metabolite signals. In the Principal Components Analysis (PCA) scores plot the dilution factor here, indicated by colour, the clustering of the samples from the five infants studied, which are indicated by different symbols, was affected (Figure 3-3a). There is a larger observable difference between individuals when the samples were more dilute as this was the biggest effect causing the separation in the model.

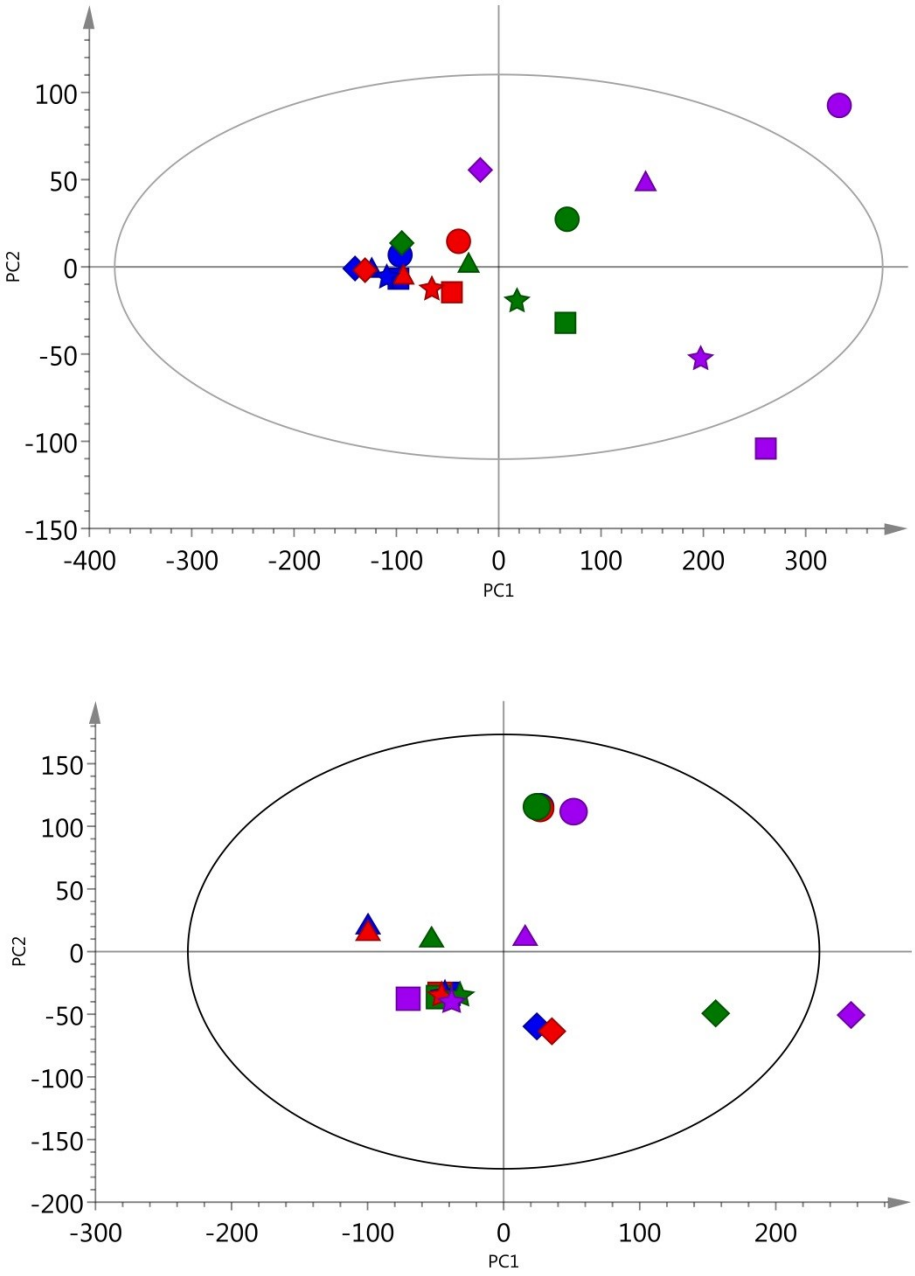


Figure 3-3 a) PCA of dilution series: R^2X =0.985 (First component accounts for 84% of the total variance), $Q^2X(cum)$ =0.941, b) PCA of normalised diluted infant samples: R^2X =0.5 (First component accounts for 32% of the total variance), $Q^2X(cum)$ =0.29 – colours represent dilution series, i) Blue, 520 µl, ii) Red, 270 µl, iii) Green, 135µl, iv) Purple, 68µl. Shapes represent different individual infants

Once normalisation was applied, the effect of dilution (Figure 3b) is reduced and clear inter-individual differences between the infants become apparent at all dilution factors, with the exception of two infants whose profiles were largely superimposed in the first two components (as indicated by the squares and stars) showing that the effect of dilution on the spectroscopic profile can be largely resolved with spectral pre-processing. For one infant represented by diamond symbols, the effect of urinary dilution was still clearly visible. This particular infant had the lowest creatinine to noise ratio, suggesting that the urine sample from this individual was the most dilute to begin with (Figure 3-4). From Figure 3-3b, it is clear that if samples are already dilute, dilutions less than half the original volume would be the most appropriate dilution factor for the best recovery of metabolic data. It is critical that all samples in a sample set are treated to the same dilution factor to gain reliability in the data.

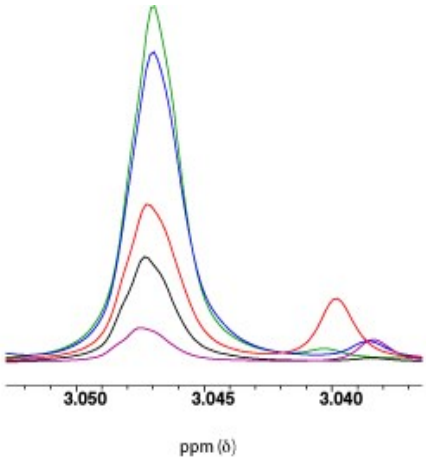


Figure 3-4. ^{1}H NMR spectra of creatinine (δ3.05) in urine from 5 infants (different colours) to determine original concentration of urine samples.

3.4.1.2. Plasma Dilution Series

When infants are born they are subject to a number of different screening tests using their blood (Pitt 2010). A small quantity of blood is obtained on a specific filter paper, and is tested for diseases including Sickle Cell Disease and Cystic Fibrosis. The spot of blood is punched out of the filter card and eluted in buffer for analysis; this is believed to be the equivalent of 3-5µl/500ml concentration (Mei et al 2001). It is necessary to limit the amount of blood drawn from babies, particularly in the

case of preterm infants. To utilise this resource, investigation of detection limits of plasma was conducted to evaluate the use of limited amounts of plasma for metabolic profiling by ^{1}H NMR spectroscopy.

Plasma gives us a metabolic picture of the circulating metabolites from areas around the body for a given moment in time. The composition of plasma is very diverse and includes both low weight metabolites and macromolecules such as lipids (Beckonert et al 2007). It is intrinsically sparser, in terms of small molecules, compared to urine and faeces.

The plasma ^{1}H NMR spectra obtained from an individual sample, with each spectrum corresponding to a different dilution factor, is shown in Figure 3-5. Samples were subject to both the same number of NMR scans as well as an increasing number of scans to adjust for the dilution series. To adjust for this difference in scans and therefore intensity of signals, samples were scaled to glucose (δ5.23). As the number of scans increased this adjusted the spectrum of all of the samples.

These results indicate that after adjustment for the number of scans, to account for differences in signal to noise, there was little visual effect of dilution on the NMR spectral signal intensities (Figure 3-5b). As mentioned previously in the urine limits results, a difference seen in the dilution samples is the peak at δ1.25, which we have identified as the contaminant tert-butanol, increases as dilutions increase. Further to this, there is a peak at δ3.38, which increases as the samples become more dilute. This unknown metabolite is only present in plasma samples and suggest this metabolite could have been protein bound in the less dilute samples.

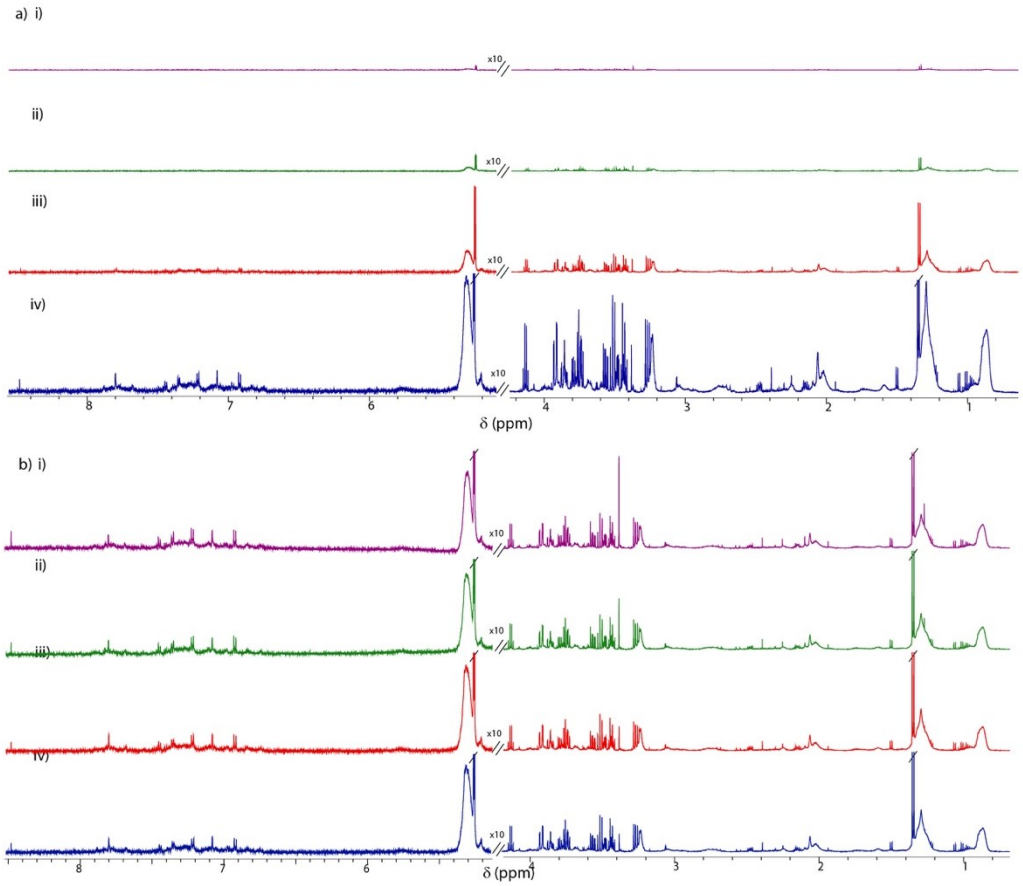


Figure 3-5 a) ^{1}H NMR CPMG plasma spectra from one infant with dilution series in effect with the same number of scans. b) CPMG ^{1}H NMR plasma spectra from one infant with dilution series with increasing number of scans as dilution increases. Dilution series: i)38µl, ii) 75µl, iii)150 µl, iv) 300 µl

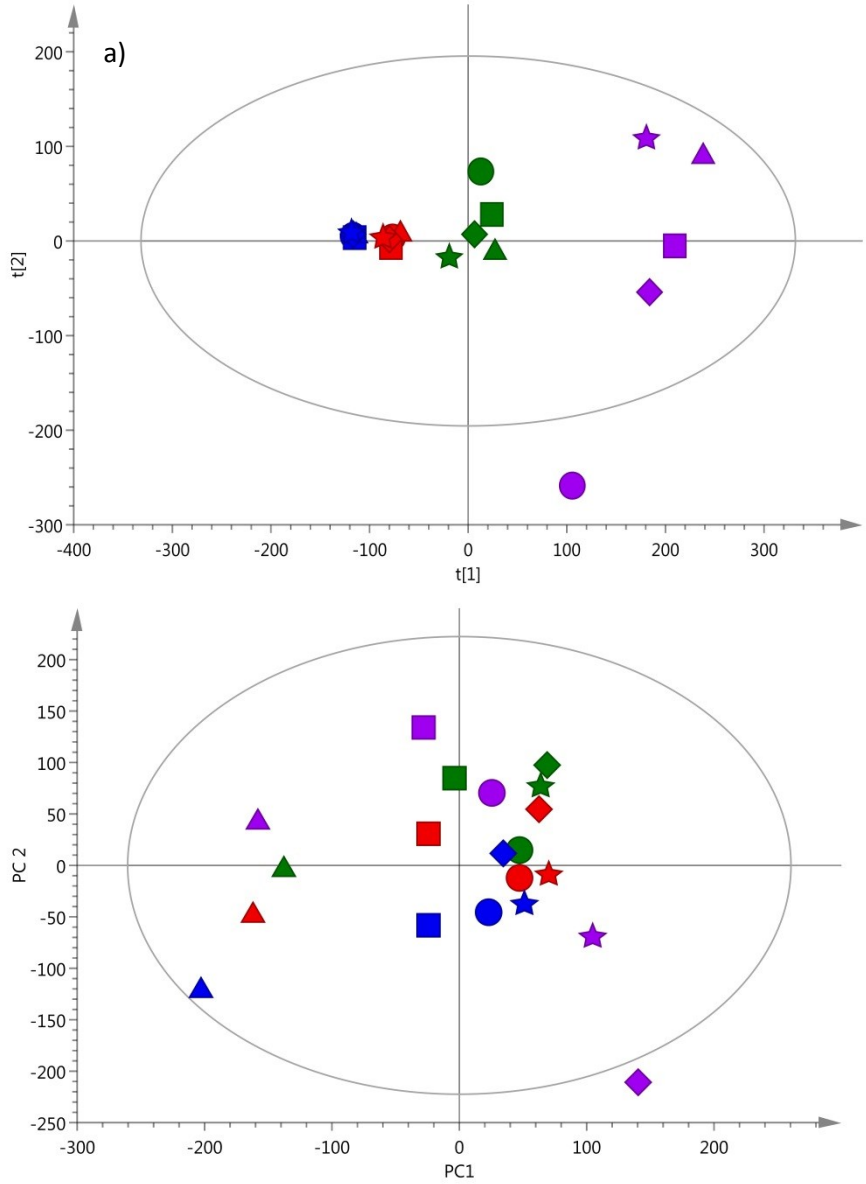


Figure 3-6 a) PCA of ^{1}H NMR CPMG plasma dilution series: R^2X=0.67, Q^2X(cum) =0.578, b) PCA of ^{1}H NMR CPMG normalised plasma diluted infant samples: R^2X =0.55, Q^2X(cum) =0.326 – colours represent dilution series, i) Blue, 300 µl, ii) Red, 150 µl, iii) Green, 75µl, iv) Purple, 38µl. Shapes represent different individual infants

The PCA scores plot (Figure 3-6a) shows how the dilution series can affect the plasma samples, even after an increase of scans to accommodate the dilution effect are introduced. Similar to the urine, the largest observable cause of variation in the data was caused by the dilution of the samples. From the raw spectral data, we have confirmed that there are some contaminants within the samples, which are at a higher concentration in the more dilute samples. Furthermore, as the dilution series increased it was harder to suppress the water signal in the spectra; the difficulty of suppressing the water strongly in the more dilute samples influenced the model in this experiment greatly.

By normalising the data, the effect of dilution on the model is greatly reduced (Figure 3-6b), and clear inter-individual differences can be recognised. This observed effect is not as strong in the plasma compared to the urine; four of the individuals have very similar plasma metabolic profiles where one individual (shown as triangles) separates in the model. When looking back at the raw data, this individual has higher relative concentration of lipids in their plasma when compared to the other participants, which explains this difference.

3.4.1.3. Faecal Water Dilution Series

Faecal water is extracted from faecal samples in a number of different ways and the optimisation of this technique will be discussed later in this chapter; section 4.4. The first point of focus will be on chemical detection limits with respect to this biofluid. Faecal water gives us an insight into not only the metabolites from our diet and biochemical events, but also from the microbiome, residing within the gastrointestinal tract (Saric et al., 2007; Jacobs et al., 2008). Thus, it enables us to study the interaction between the gut microbiome and host more closely (Nicholson et al., 2012). It is increasingly apparent that this knowledge is especially important in early life as the gur microbiome are involved in infant development, including the immune system.

The faecal water ^{1}H NMR spectra obtained from an individual sample defined by a dilution series, with each spectrum corresponding to a different dilution factor, is shown in Figure 3-7. Samples were subject to both the same number of NMR scans as well as increasing the number of scans to correspond to the adjustment factor necessary to accommodate the dilution series.

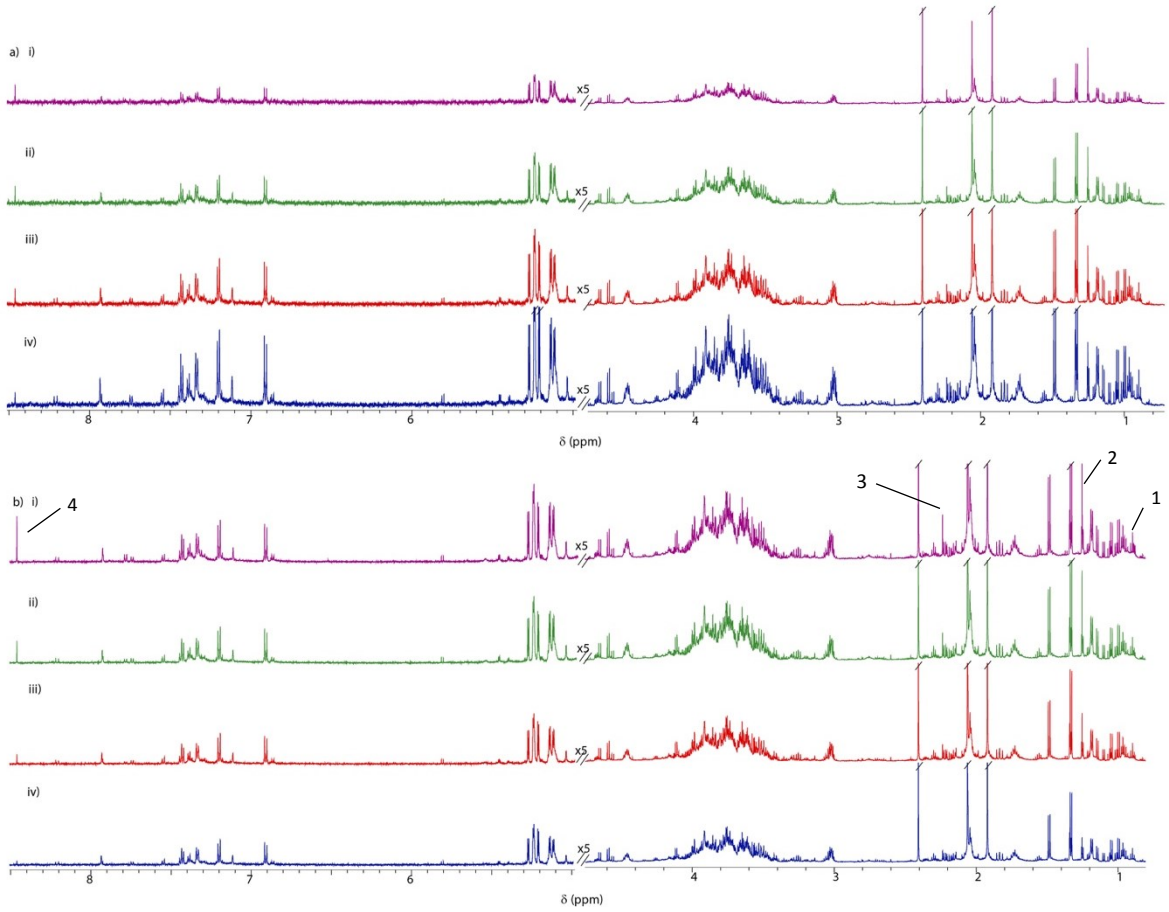


Figure 3-7 a) 1D ^{1}H NMR faecal water spectra from one infant with dilution series in effect with the same number of scans. b) 1D ^{1}H NMR faecal water spectra from one infant with dilution series with increasing number of scans as dilution increases. Dilution series: i)25mg, ii) 50mg, iii)100 mg, iv) 200 mg. Key: 1. Iso-butanol (δ0.88), 2. Tert-butanol (δ1.25), 3. Acetone (δ2.23) and 4. Formate (δ8.46)

As mentioned previously, a difference seen in the dilution samples is the peak at δ1.25, which was identified as the contaminant tert-butanol, which increased as the dilution factor increases. There is an increase signal at δ0.88, δ2.23 and δ8.46, which have been identified as iso-butanol, acetone and

formate respectively. These metabolites are also known to be contaminants of biofluids for metabolic phenotyping. Formate, which can be endogenous, is also known to be related to the gut microbiome, this co-metabolite may indicate that there was microbial activity during the NMR run; especially as the increase in scans causes the experiment to be much longer in length.

The PCA scores plots (Figure 3-8a) shows how the dilution series can affect the faecal samples, even after an increase of scans are introduced. From what we can observe, similar to the urine, the largest observable variation in the data can be attributed to the dilution of the samples. From the raw spectral data, we have confirmed that there are some contaminants within the samples are at a higher concentration in the more dilute samples.

By normalising the data, the effect of dilution on the model is greatly reduced (Figure 3-8b), and clear inter-individual differences between individuals can be recognised. Normalisation can clearly be seen in the individuals depicted as squares as their scores cluster together, indicating that their ^{1}H NMR spectra are extremely similar. Interestingly, individuals depicted as diamonds and stars have very similar faecal metabolic profiles to each other.

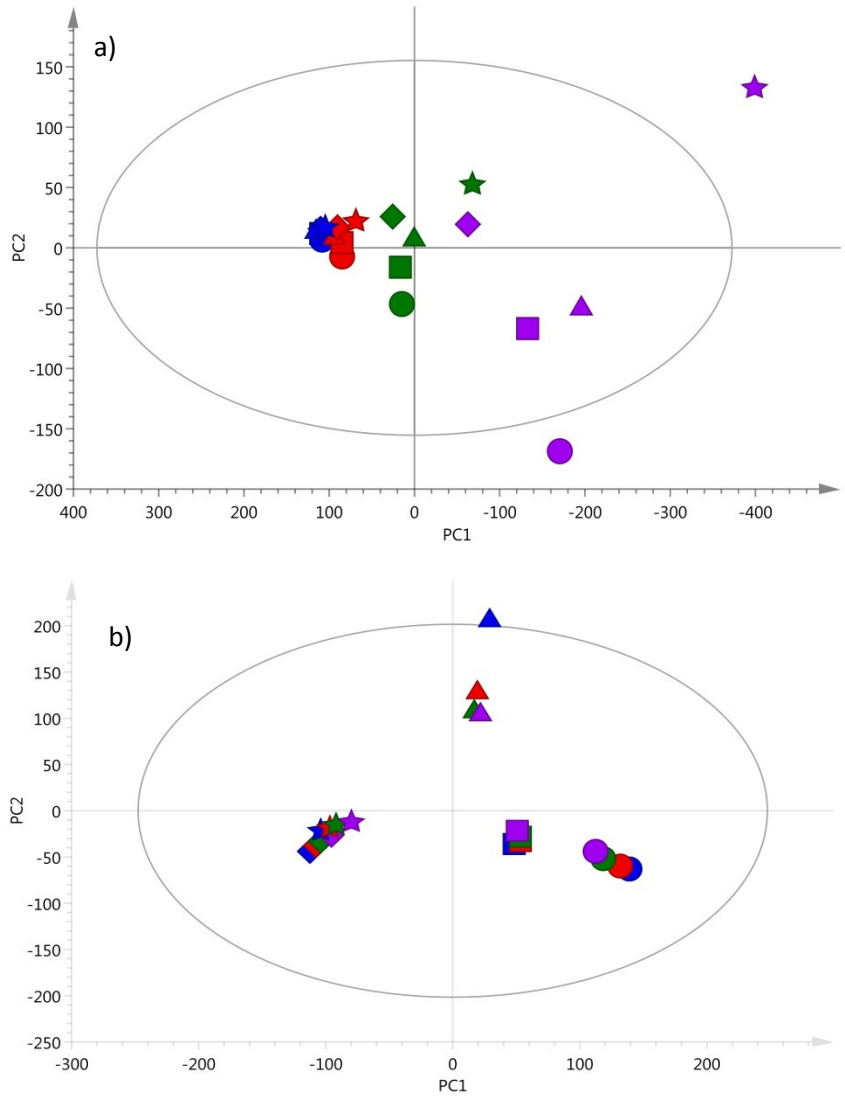


Figure 3-8 a) PCA of dilution series: $R^2X = 0.78$, $Q^2X(cum) = 0.562$, b) PCA of normalised diluted infant faecal samples: $R^2X = 0.488$, $Q^2X(cum) = 0.31$ – colours represent dilution series, i) Blue, 200mg, ii) Red, 100mg, iii) Green, 50mg, iv) Purple, 25mg. Shapes represent different individual infants

3.4.2. AGE VARIABILITY IN INFANT URINE PROFILES

The urinary ^{1}H NMR spectral profiles of infants at the time of birth are consequently very similar to urinary samples collected from the mother at the same time (Figure 3-9a). The similarities are particularly clear in the aromatic region of the NMR spectra, where there is higher relative concentration of hippurate (δ3.95) and phenylacetylglutamine (δ2.14, δ2.46, δ3.77, δ7.35) in the infant sample obtained at birth than one would expect, given that many papers report an absence of signals in the spectral window of infant urine where phenolic and other aromatic molecules would be expected to arise. From the literature, microbiome colonisation of the infant gut largely begins at birth and that over a period of time as the symbiotic gut bacteria evolve, there is an increase in molecules such as hippurate, phenylacetylglutamine (PAG), indoles and methylamines (Holmes et al 2011). Here the urine sample obtained immediately at birth, hippurate (δ3.95), 4-cresylsulfate (δ2.22) and PAG (δ2.14, δ2.46, δ3.77, δ7.35) are found in the infant urine in similar concentrations to the maternal urine. However, some differences between maternal and infant urine composition are visible at this initial sampling point, including the presence of higher concentrations of trimethylamine (TMA- δ2.28) and formate (δ8.45) with lower concentrations of methylhistidine (δ6.93, δ7.69).

After two weeks (Figure 3-9(c)), the infant's urine spectrum has become more dilute compared to the mother's urine. The most striking difference between these spectra is that there is much less creatinine and no hippurate with a reduction in the excretion of most aromatic compounds in the week two sample from the infant compared to the birth sample. This reinforces the observation of the close relationship between the mother and foetus exchange of nutrients and waste products during pregnancy and immediately postpartum which dissipates over the first few days. At this time, betaine (δ3.25, δ3.89) is also found in higher concentrations in the neonatal urine compared to the maternal or infant urine samples taken later postpartum.

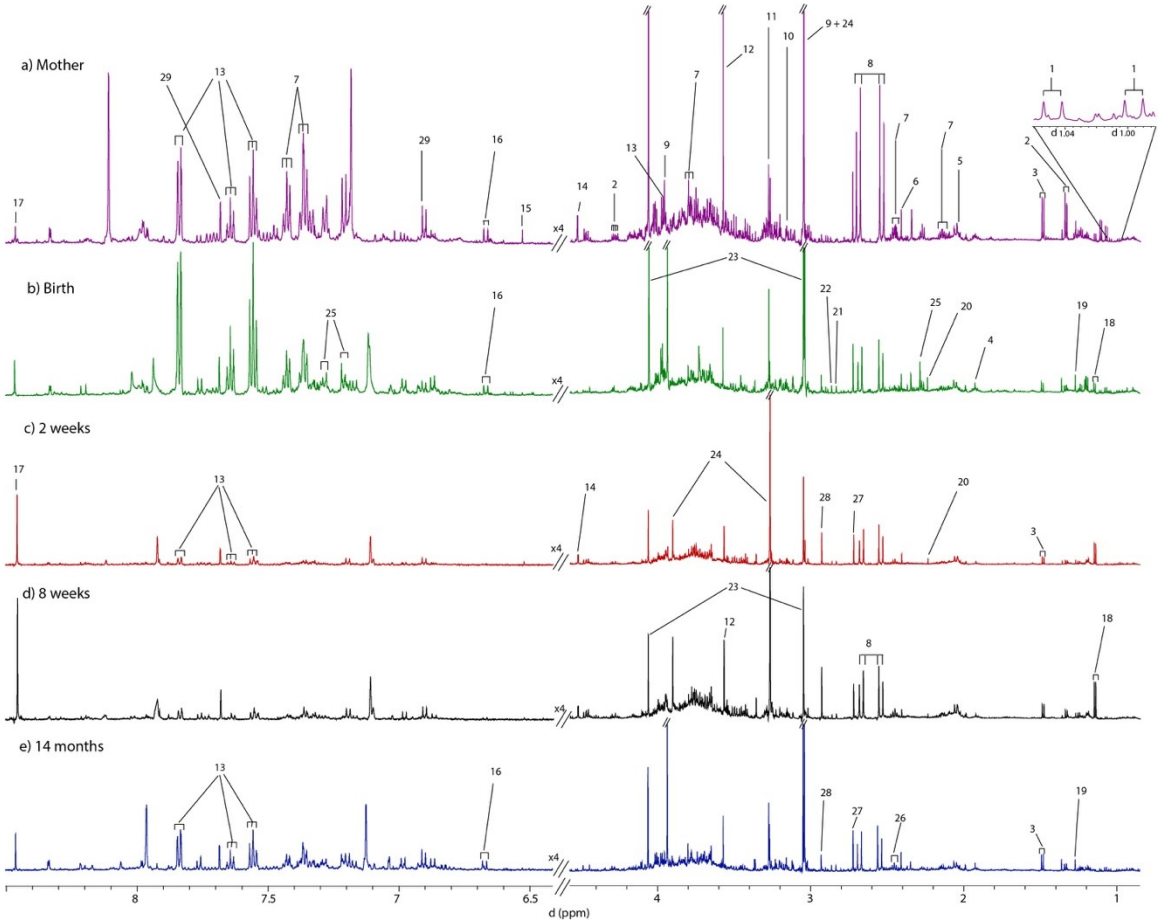


Figure 3-9: Mother – Baby Urine match 1D ^{1}H NMR Spectra showing the age variability from birth until 14 months postpartum. a) Mother b) Birth c) 2 weeks d) 8 weeks and e)14 months. 1. Valine (δ0.99 $–CH_3$; δ 1.04 $–CH_3$); 2. Lactate (δ1.33 $-CH_3$; 4.11 -CH); 3. Alanine (δ1.47 $-CH_3$); 4. Acetate (δ 1.92 $–CH_3$); 5. Uromodulin (δ2.06 $–CH_2$) 6. Succinate (δ2.42 $-CH_2$); 7. Phenylacetylglutamine (PAG) (δ 2.14 $–CH_2$; δ2.46 $-CH_2$; δ3.77 $-CH_2$; δ7.35 $(–CH)_3$); 8. Citrate (δ2.55 $-CH_2$; δ2.66 –CH); 9. Creatine (δ3.02 $-CH_3$ (overlapped with creatinine); δ3.92 $-CH_2$) 10. Proline betaine (δ3.11 $-CH_3$; δ3.30 $-CH_3$); 11. Trimethylamine-N-oxide (δ3.25 $-CH_3$); 12. Glycine (δ3.55 $-CH_2$); 13. Hippurate (δ3.95 -CH); 14. N-methylnicotinamide (δ4.48 -CH); 15. Fumarate (δ6.53 $-CH_3$); 16. Urocanic acid (δ6.65); 17. Formate (δ8.45 -CH); 18. Propylene glycol (δ1.14 $-CH_3$); 19. Tert-butanol (δ1.25 $-CH_3$); 20. Acetone (δ2.24 $-CH_3$); 21. Methylguanidine* (δ2.23 $-CH_3$); 22. Trimethylamine (δ2.28 $-CH_3$); 23. Creatinine (δ3.05 $-CH_3$; δ4.06 $-CH_2$); 24. Betaine (δ3.25 $(-CH_3)_3$; δ3.89 $-CH_2$); 25. 4-cresol sulphate (δ2.22 $-CH_3$; δ7.11 –CH; δ7.18 -CH); 26. 2-oxoglutarate (δ2.42 -CH); 27. Dimethylamine (δ2.72 $-CH_3$); 28. Dimethylglycine (δ2.93 $-CH_3$); 29. 3-methyl-histidine (δ6.93 -CH; δ7.69 -CH). *tentative assignment

At 8 weeks (Figure 3-9(d)), the infant's urine spectrum is still dilute compared to the mother's urine and very similar to the sample taken after two weeks. Creatinine (δ3.05, δ4.06) levels have increased slightly and betaine (δ3.25, δ3.89) is still high in this early sample. At this age, infants are being fed a milk based diet (whether it be breast milk or formula milk) and co-metabolites produced from gut microorganisms in the aromatic region are low except for formate (δ8.45). Glycine (δ3.55), an essential amino acid, is also elevated in this sample which could indicate an increased dietary consumption.

At 14 months (Figure 3-9(e)), the infant will have grown considerably and be weaned onto solid foods as well as have a larger bacterial diversity in their gut microbiome, which may contribute to their metabolism and therefore their urine metabolic profile. The ^{1}H NMR spectra taken at 14 months postpartum shows a higher concentration of creatinine (δ3.05, δ4.06), which is related to muscle mass, as well as other metabolites (e.g hippurate and PAG) found in the aromatic region, which are co-metabolites produced from gut microorganisms.

3.5. NAPPY AND COTTON WOOL CONTAMINATION ON INFANT URINE METABOLIC PROFILES

3.5.1. Cotton wool and nappy contamination

Visual comparison of the ^{1}H NMR spectra, obtained from blank water infused cotton wool samples, were assessed for chemical contaminants (Figure 3-10). From cotton wool, the major contaminants were identified as acetate (δ1.92 (-CH_3)), acetone (δ2.23 (-CH_3)), formate (δ 8.46 (-CH)), citrate (δ2.55(-CH_2), δ2.66(-CH_2)) and propylene glycol (δ1.14 (-CH_3), δ3.4 (–CH), δ3.6 (–CH), δ3.9, (-CH_2)), tert-butanol (δ1.25 (CH_3)), iso-butanol (δ0.88 (-CH_3)) as well as a peak at δ6.02 (currently unidentified). Metabolite identification was assisted by our in-house database, together with the aid of 2-D ^{1}H-^{1}H COSY; a dotted line representing propylene glycol identification in Figure 3-10b.

Furthermore, despite the introduction of contaminants from the cotton wool, the urine profiles were highly similar and a change in citrate (δ2.55, δ2.66) concentration was the only visual difference in endogenous compound concentration after urine had been extracted from the cotton wool, which suggests that citrate may be partially retained in the cotton wool. Furthermore, the citrate signals became broader after treatment with cotton wool which also suggests that there was binding of this metabolite (Figure 3-10(c)).

Different cotton wool brands contained variable concentrations of these metabolites; for example, it can clearly be seen there is a lower concentration of formate in the 'Boots' own brand cotton wool compared to the other cotton wool brands (Figure 3-11). However, this brand has a much higher concentration of acetate compared to the cotton wool in other brands tested. Whereas all cotton wool samples emitted the same amount of acetone into the urine, the 'Sainsburys' own brand cotton wool leached higher amounts of tert-butanol. However, overall, the qualitative presence of the metabolites was consistent between all the brands of cotton wool.

Contaminants from nappies included propylene glycol (δ1.14 ($-CH_3$), δ3.4 (–CH), δ3.6 (–CH), δ3.9 ($-CH_2$)), N, N-dimethylacetamide (δ2.1 ($-CH_3$), 2.9 ($-CH_3$), 3.1 ($-CH_3$)) ethylene glycol (δ3.7 (-CH)), tert-butanol (δ1.25 ($-CH_3$)), isobutanol (δ0.88 (CH_3)); as well as acrylic acid (δ5.76, δ6.02, δ6.11) and propionate (δ2.19 ($-CH_2$), δ1.06 ($-CH_3$) (Figure 3-10a). Similar to what was observed with the different brands of cotton wool, different amounts of the contaminants were seen between the different brands; including higher concentrations of acetone in Libero nappies and higher amounts of tert-butanol in the Pampers brand (Figure 3-12). Furthermore, the Tesco nappy own brand, had higher amounts of the unknown metabolite at δ6.02. However, acrylic acid was only seen as a contaminant in the Pampers brand of nappy and not within the other two brands tested in this investigation. A nappy's absorbent core is predominantly derived from a polymer made from very small particles of acrylic acid derivatives, giving its highly absorbent nature and seen in top market brands of nappy (e.g. Pampers).

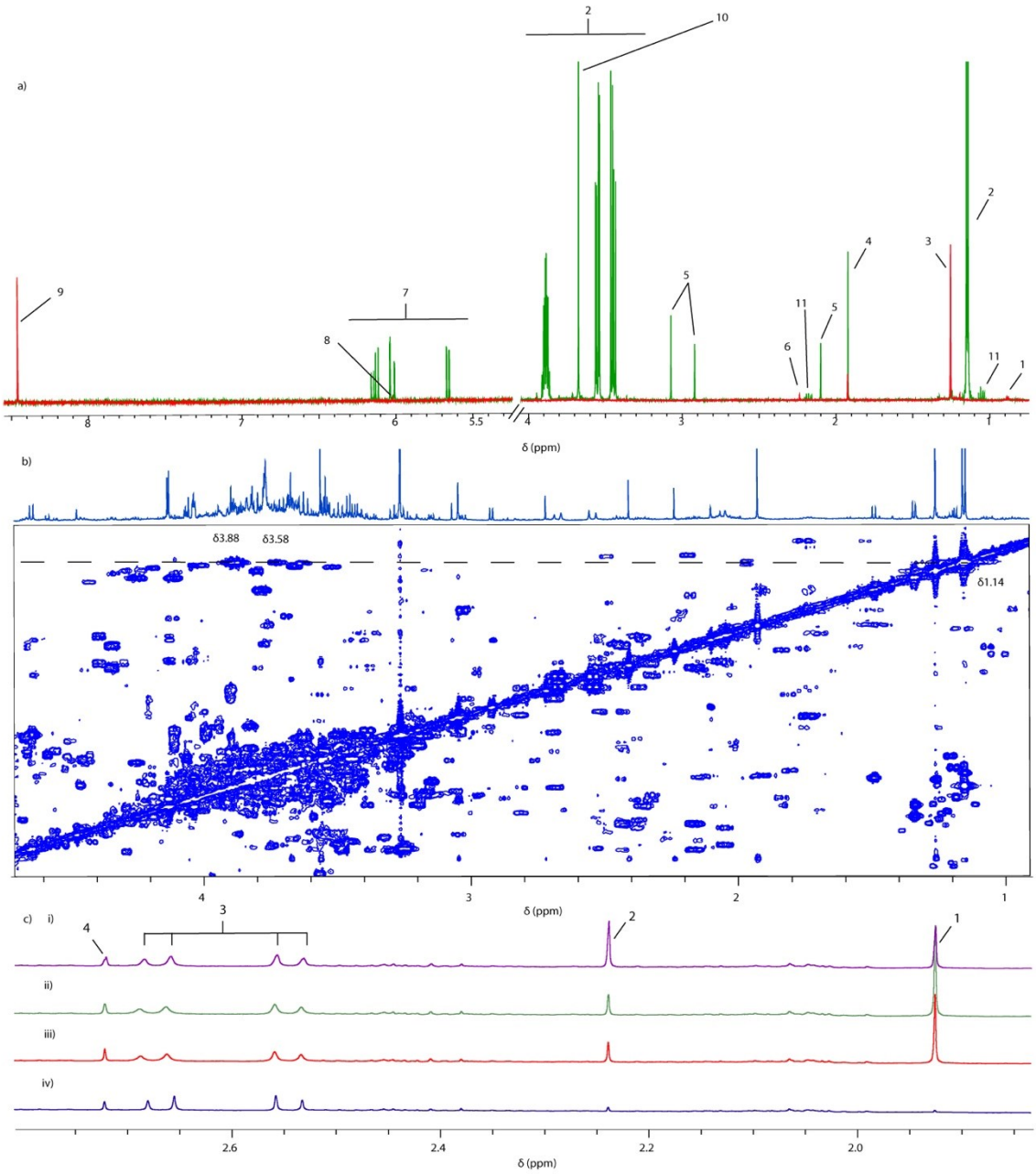


Figure 3-10: a) 1D ^{1}H NMR spectra from cotton wool (red spectrum) and diaper blanks (green spectrum), Key: 1. Isobutanol (δ0.88 -CH_3); 2. Propylene glycol (δ1.14 -CH_3; 3.43 -CH_2; 3.53 -CH_2; 3.87 -CH); 3. Tert-butanol (δ1.25 -CH_3); 4. Acetate (δ1.92 -CH_3); 5. N, N-dimethylacetamide (δ2.1 -CH_3; δ2.9 -CH_3; δ3.1 -CH_3); 6. Acetone (δ2.23 -CH_3); 7. Acrylic Acid (δ5.76 –CH; δ6.02 –CH; δ6.11 -CH); 8. Unknown metabolite (δ6.02); 9. Formate (δ8.46 -CH); 10. Ethylene glycol (δ3.7 –CH); 11. Propionate (δ2.19 -CH_2; δ1.06 -CH_3); b) 2D COSY ^{1}H NMR spectra, dotted line depicting Propylene glycol (δ1.14 -CH_3; 3.43 -CH_2; 3.87 -CH); c) 1D ^{1}H NMR partial spectra (δ1.85 - δ2.75) from cotton wool which have been subject to freeze-thaw cycles; i) urine extracted from cotton wool and subject to one freeze thaw cycle; ii) urine extracted from cotton wool after one freeze thaw cycle; iii) urine extracted from cotton wool and acquired immediately; iv) control blank urine. Key: 1. Acetate (δ1.92), 2. Acetone (δ2.24), 3. Citrate (δ2.55, δ2.66), and 4. Dimethylamine (δ2.72).

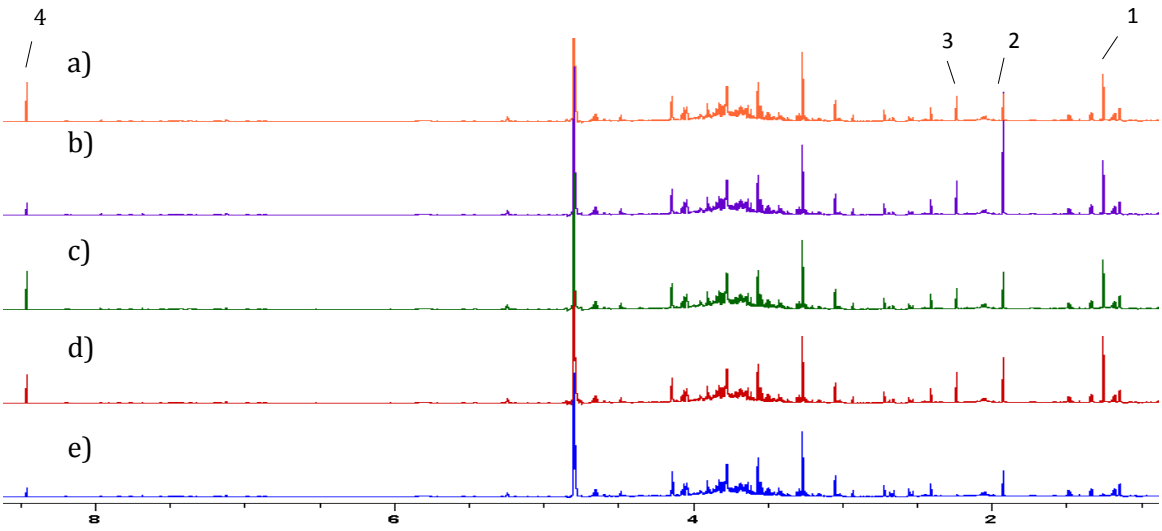


Figure 3-11. 1D ^{1}H NMR spectra of a QC urine sample and from the same urine subject to processing through five different cotton wool brands. a) Hospital brand – Robinson's, b) Boots c) Tesco, d) Sainsburys, and e) original urine sample without cotton wool contamination. 1. Tert-butanol (δ1.25), 2. Acetate (δ1.92), 3. Acetone (δ2.23) and 4. Formate (δ8.46).

Nappies are made of highly absorbent polymer structures including polypropylene and acrylic acid derivatives to ensure as much liquid is retained in them as possible. From Figure 3-12, the nappy and cotton wool profile are comparably different, with nappies leaching more polymer metabolites into the urine and causing higher amounts of contamination than cotton wool. The high concentration of PEG in these samples masks the urinary biological profile making analysis very difficult.

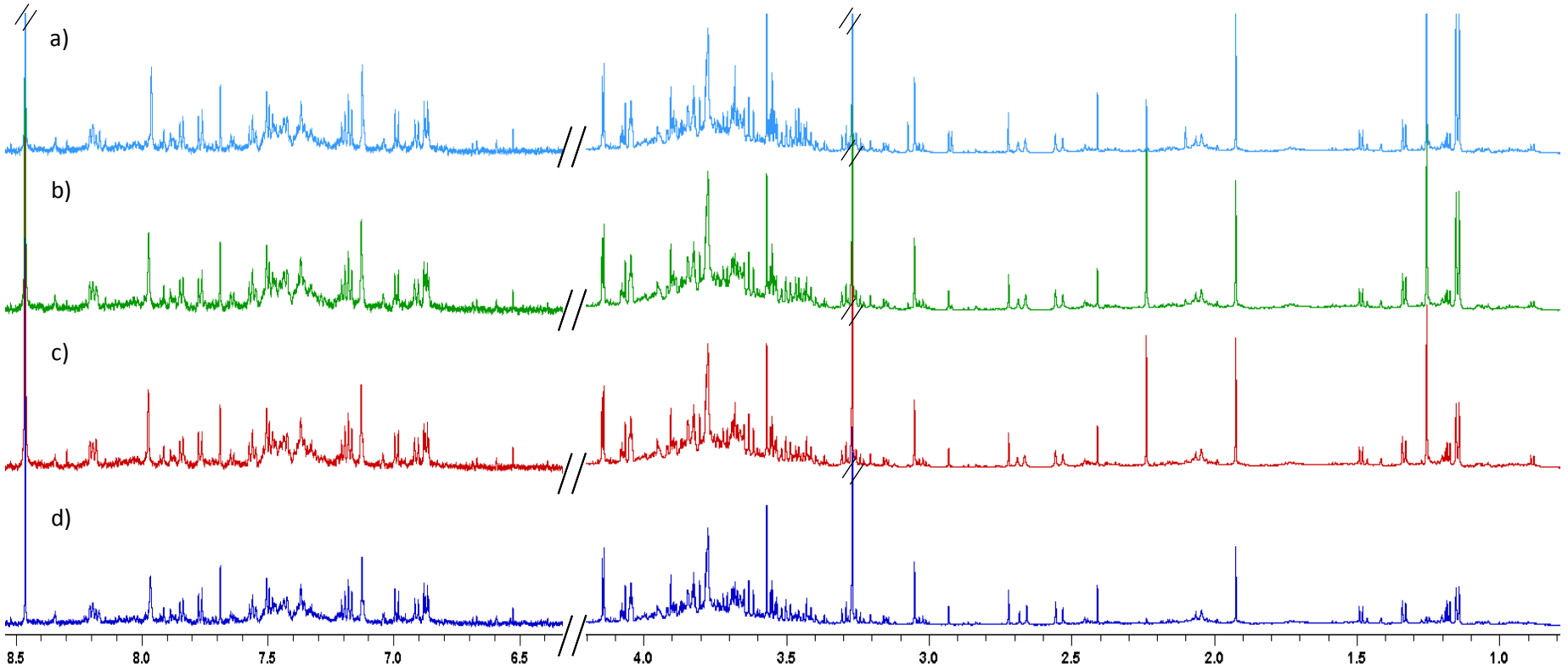


Figure 3-12. 1D ^{1}H NMR spectra of a QC urine sample and from the same urine subject to processing through three different nappy brands. a) Pampers nappy, b) Libero nappy, c) Tesco nappy, d) blank urine.

Similar results were obtained for the UPLC-MS analysis, with the addition of high concentration of polyethyleneglycol (PEG) and other polymers (Figure 3-13). Cotton wool is a natural fibrous material and a high amount of PEG was observed in the sample, especially when running HILIC chromatography in positive ionisation mode.

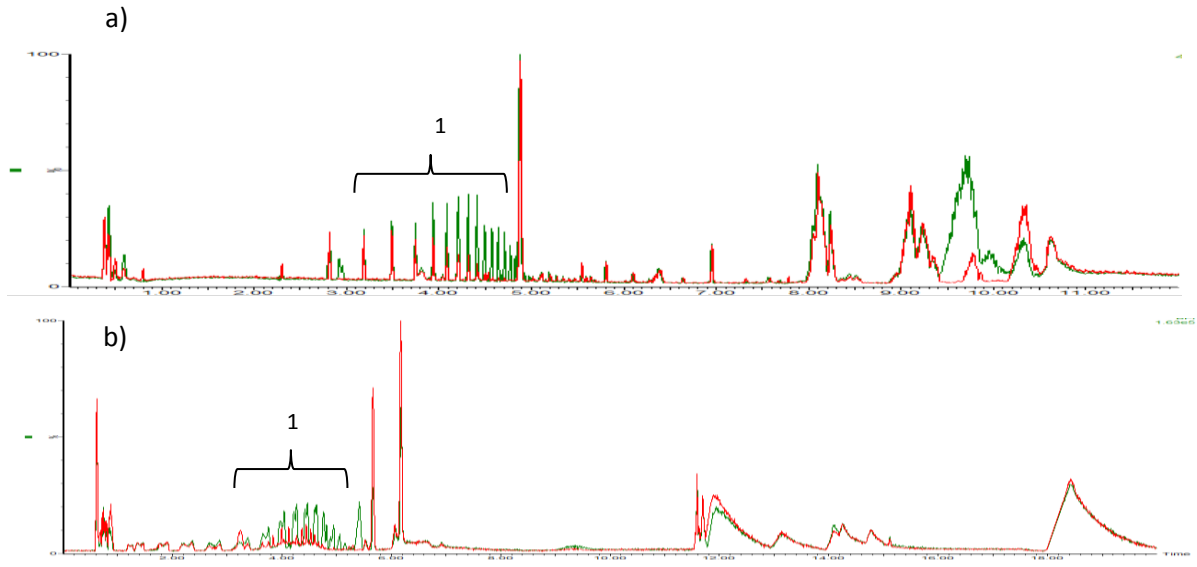


Figure 3-13 a) Base Peak intensity (BPI) chromatogram from RP UPLC-MS (ESI+) - cotton wool (red) and diaper blanks (green) 1. Polyethylene glycol (PEG) b) Base Peak intensity (BPI) chromatogram from HILIC UPLC-MS (ESI+) - cotton wool (red) and diaper blanks (green) 1. Polyethylene glycol (PEG)

3.5.2. Contamination from cotton wool preparation

To extract the urine from the cotton wool for analysis, the cotton wool was either squeezed manually or centrifuged using the apparatus depicted in supplementary material (Figure S2). The cotton wool, which had been squeezed to extract the urine, had a much higher concentration of acetone (δ2.24), tert-butanol (δ1.25), and iso-butanol (δ0.88) (Figure 3-14) than the urine obtained via centrifugation. This suggests that manually squeezing the cotton wool results in greater destruction of the cotton wool fibres and therefore releases more contaminants into the urine

sample. This further underlines the need for better defined standard operating procedures and materials within this research area.

The effect of freezing the urine sample in the cotton wool prior to extraction was further assessed and compared with obtaining urine fresh from the cotton wool ball. No obvious differences in the level of chemical contamination or endogenous profile were found. Therefore, freezing the cotton wool directly after collection at -80 °C is an acceptable practice when collecting urine from infants.

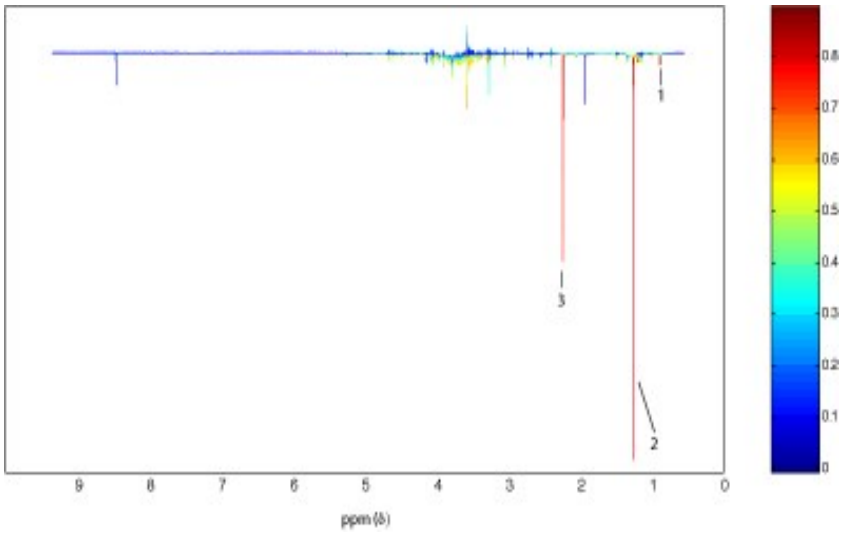


Figure 3-14. Correlation coefficient plot by two groups; cotton wool squeezed (negative) vs centrifuged (positive) coloured by colour intensity. This plot highlights the discriminatory metabolites for these two groups. 1) isobutanol (δ0.88), 2) tert-butanol (δ1.25), 3) acetone (δ2.23)

3.6. OPTIMISATION OF FAECAL WATER EXTRACTIONS FOR ^{1}H NMR SPECTROSCOPY

3.6.1. Freeze-thawing of faecal material

Aliquots from a well-homogenised individual sample were taken and were subjected to different thawing cycles to establish the optimum method. By thawing the sample at least once, the concentration of metabolites increases and the highest concentration of metabolites can be seen after two freeze-thaw cycles (Figure 3-15). Interestingly, after three cycles of freeze thawing, the sample had a lower concentration of metabolites, which could be due to degradation of the sample in the last freeze thaw cycle.

3.6.2. Homogenisation techniques

In this investigation, four different extraction processes were used; 1) vortexing, 2) bead beating, 3) pestle homogenisation in microcentrifuge tube and 4) liquid nitrogen homogenisation using pestle and mortar. Other recommended extraction processes, including freeze drying and tissue lysing, were not included in this investigation, as they required excessive disruption of the sample including stabilising with acid and alkaline solutions and therefore not suitable for high throughput investigations.

In Figure 3-16, the ^{1}H NMR spectrum shows that the bead beating technique was able to extract the highest relative concentration of metabolites. For bead beating (Figure 3-16d), the added density to the sample in the form of the beads and the uniformity of the beating technique improves the homogenisation. Examination of acquired raw spectra revealed that the same biochemical information was obtained from each of the extraction processes. The vortex technique produced a considerably lower concentration of metabolites (Figure 3-16e). Vortexing the sample to extract faecal water isn't as strong a method at homogenising the sample compared to the other methods tested here.

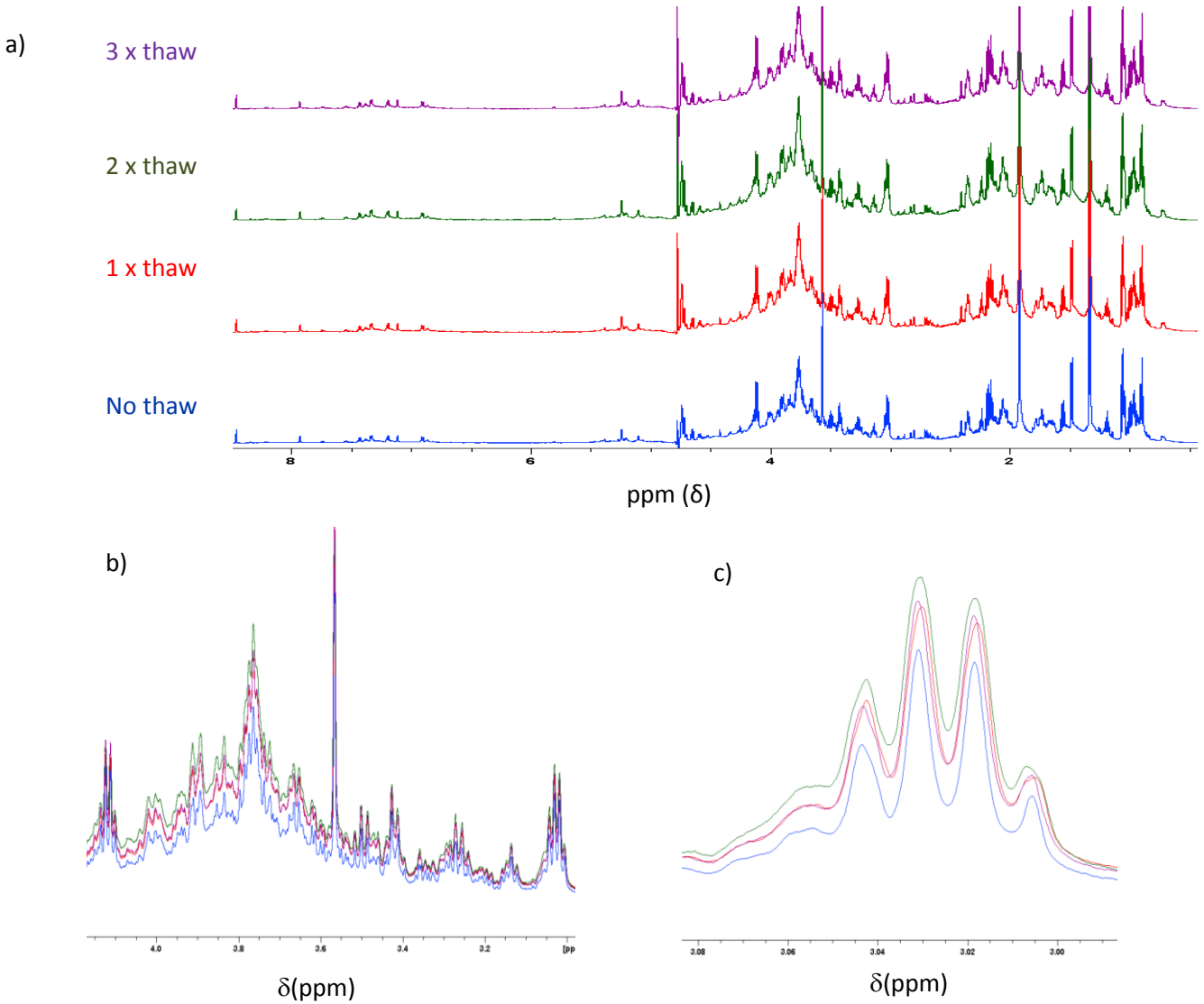


Figure 3-15: 1D ^{1}H NMR spectra of infant faecal water extractions after different freeze-thaw cycles. No thawing of sample, Freeze thawing cycles once, twice and three times. (a) Full spectra separate (b) overlaid full spectra and (c) lysine (δ3.02 (q)).

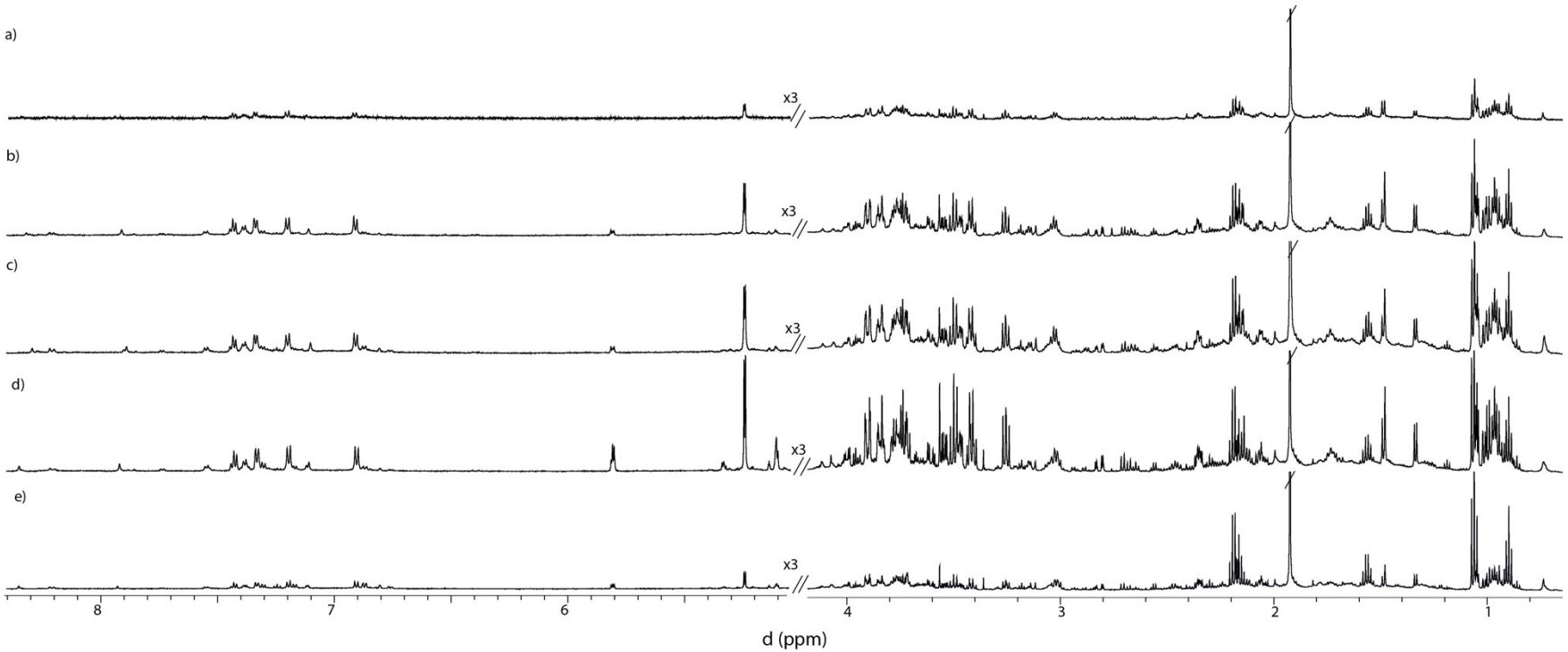


Figure 3-16: 1D ^{1}H NMR spectral data of infant faecal water extractions in buffer. Different extraction methods in Buffer at 250mg (a) Centrifuge only (b) Pestle Homogenisation (c) Liquid Nitrogen (d) Bead beating (e) Vortex

The most robust techniques for faecal water extraction are bead beating and liquid nitrogen homogenisation techniques. Spectra from the bead beating generally had a higher concentration of metabolites; this increase was especially true for carbohydrate metabolites (δ3-5). However, liquid nitrogen homogenisation spectra had higher concentrations of the metabolites butyrate (δ1.56), prolinebetaine (δ3.11) and a species of bile acid (δ0.73). There are also some shift differences in the spectra between the two techniques which could be an indication of a difference in pH. Both techniques use mechanical homogenisation, however bead beating has a higher throughput for screening processes potential.

Additionally, to these extraction processes, investigation of the extraction of faecal water from the sample without any interference except for centrifuging was performed. In Figure 3-16a, the ^{1}H NMR spectra show that by centrifuging only, information can still be gained from the sample, however the amount of signal recovery is much less in terms of concentration of metabolites compared to other extraction techniques. Therefore, a centrifugation only protocol may be an option if the sample is needed for further investigation using different analytical platforms and it would be beneficial to have no previous treatment of the sample. This, however, is not optimal and challenging for high throughput techniques.

Acquired NMR spectral data were analysed using PCA scores plot (Figure 3-17), where the samples are coloured by extraction method and revealed separation between the different extraction processes. Pestle and liquid nitrogen homogenisation appeared to be very similar as their spectral profiles were closely clustered in the scores scatter plot, whereas bead beating and vortexing techniques have the highest separation from the rest of the sample, indicating that the spectral profile of faecal water extracted, following these two techniques, was different. Reviewing the raw individual spectra, the concentration of metabolites are different, but there are no additional metabolites from the different homogenisation techniques.

Interestingly, the centrifugation only method generated very similar information from the spectra to the rest of the extraction methods and suggests that this method does not loose information to cause it to lose reproducibility.

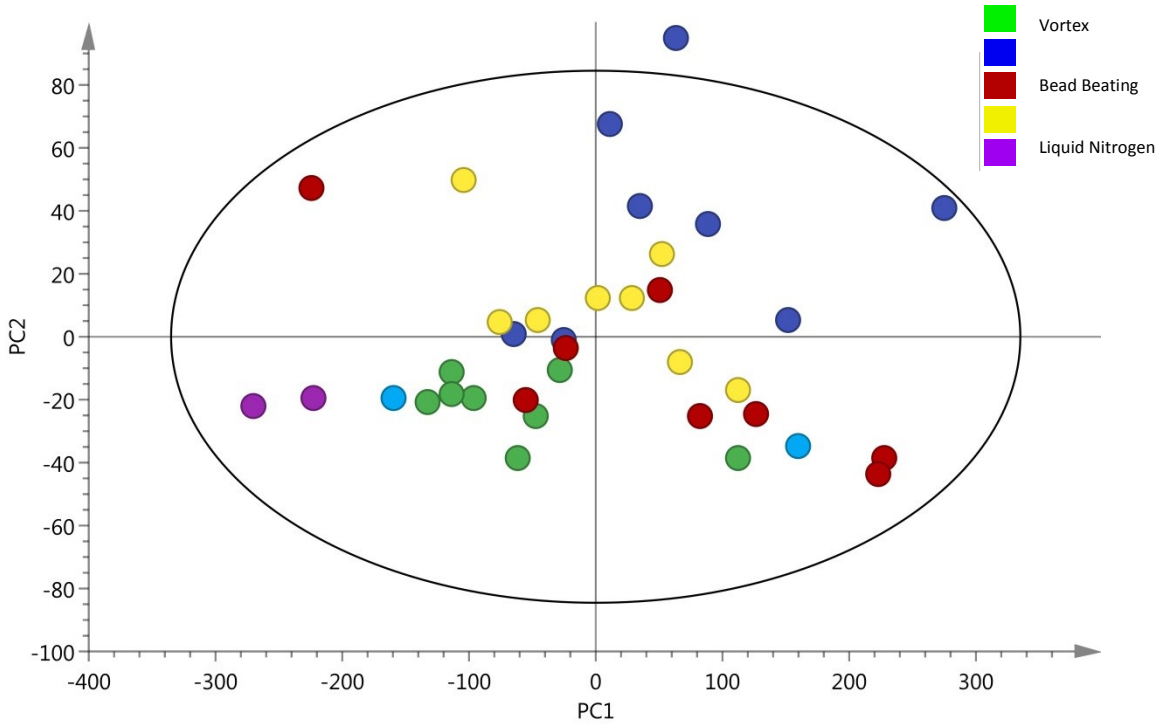


Figure 3-17: Multivariate analysis of faecal water extraction samples using PCA. $R^2(X)$ =0.80, $Q^2(cum)$ =0.562. Different colours represent different faecal water extraction techniques; Green = Vortex, Blue=Beadbeating, Red=Liquid Nitrogen, Yellow=Pestle homogenisation, Purple=Centrifugation only.

3.6.3. Solvent extraction

Whilst investigating the extraction methods, extracting into different solvents was also investigated; including deionised water, D_2O, phosphate buffer and phosphate buffered saline (PBS). Jacobs et al (2008) compared extracting faecal samples using water and methanol; they discovered that by extracting by methanol it optimised the extraction of lipids and water was optimal for extraction of

amino acids and sugars; as lipids are not the metabolite class of interest for this project methanol was not explored in this investigation.

From the raw NMR spectra (Figure 3-18), it can be seen that by using different solvents, the concentration and resolution of peaks in the NMR spectra is varied. Interestingly, phosphate buffer, which is commonly used for metabolic profiling by NMR, recovered the lowest concentration of metabolites.

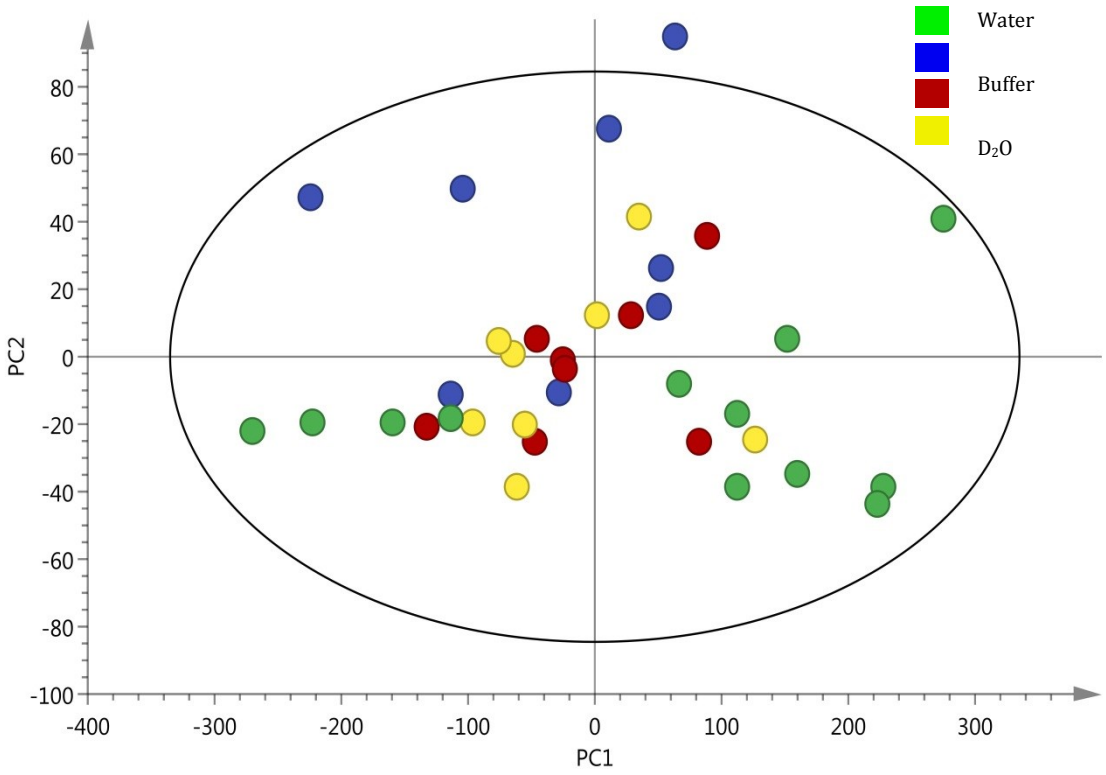


Figure 3-18: Multivariate analysis of faecal water extraction samples using PCA. $R^2(X)$ =0.80, $Q^2(cum)$ =0.56. Different colours representing different solvent extractions for faecal water: Green =water (H_2O), Blue = Phosphate buffer, Red = Deuterated water (D_2O) and Yellow = Phosphate Buffered Saline (PBS)

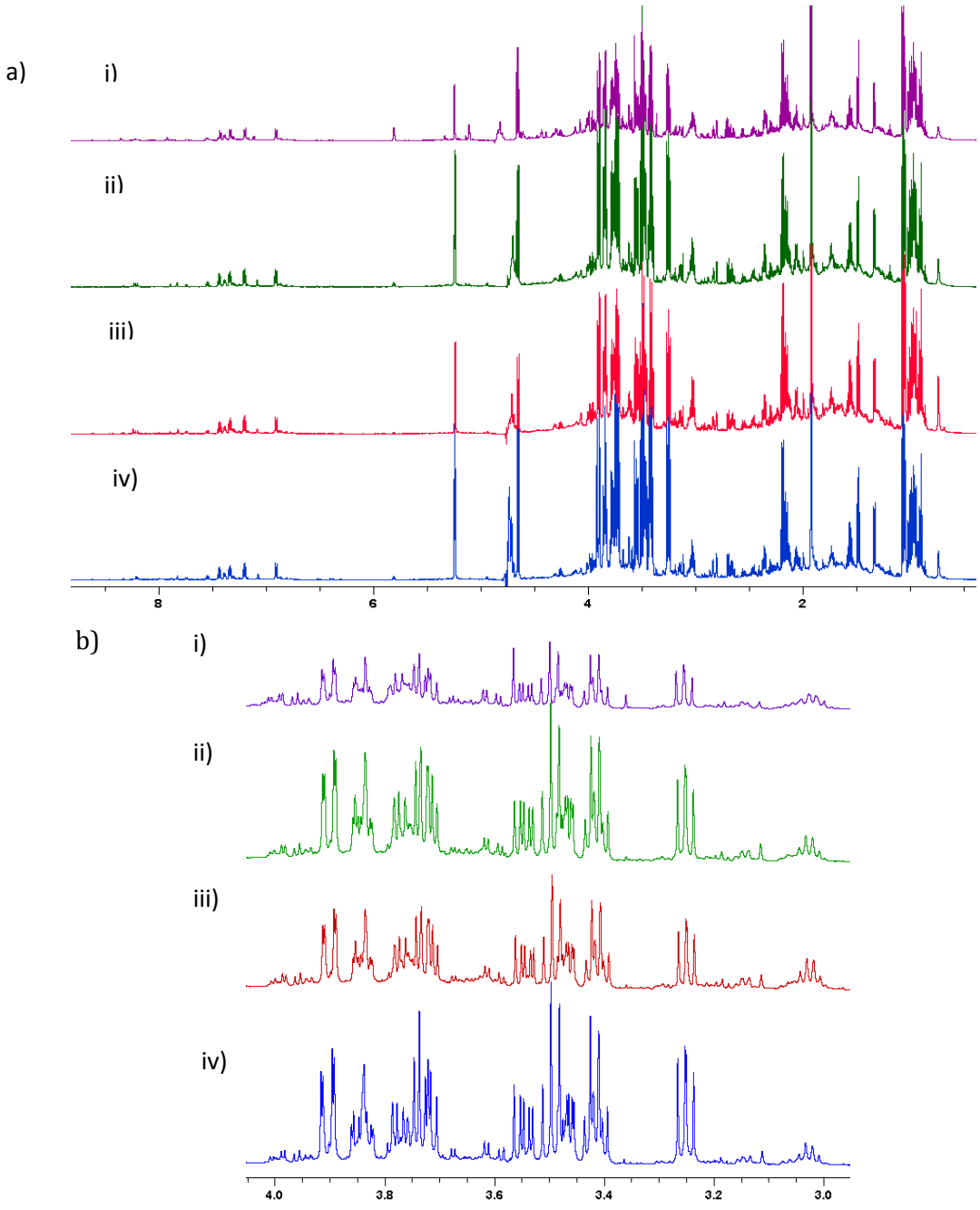


Figure 3-19: 1D ^{1}H NMR spectra of infant faecal water extractions using different solvents in bead beating extraction technique using 250mg mass; Phosphate buffer, Phosphate Buffered Saline (PBS), D_2O and Water. a) Full NMR spectra, b) partial NMR spectra δ3.0 - 4.0 - Extracted solvents: i) phosphate buffer, ii) Phosphate Buffered Saline (PBS), iii) deuterated water (D_2O) and iv)

During the faecal water extractions, it was observed that when extracting into solvents that were deuterated, there were large particles present in the sample, which floated to the top over time. This suggests that may be some larger macromolecules, which are unable to dissolve in deuterated solvents and could cause some smaller molecules to bind to them. These are thus undiscoverable by metabolic profiling and can cause technical problems with the acquisition of ^{1}H NMR data due to the lack of homogeneity of the sample; this is cause for concern when using 100% deuterated solvents for extraction purposes.

Water extracted the highest concentration of metabolites and had the best resolution in the NMR spectra. However, for dilute samples an important instrumental challenge is optimal suppression of the water peak. If the correct position of the water peak has not been offset, the method may not be amenable to automation, which is not ideal for a large sample set and high throughput experiments. However, by using a non-deuterated solvent such as water to perform the faecal water extraction, the sample is not compromised for analysis by other platforms including liquid chromatography-mass spectroscopy.

In Figure 3-19, the PCA scores plot is coloured by the different solvents that the faecal sample was extracted into; the close clustering of samples reveals that there is not a significant difference in faecal water metabolic profile, following extraction using different solvents, although extraction into buffer gave the most variable results.

3.6.4. Faecal water extraction for UPLC-MS techniques

Infant samples are particularly valuable as samples volume can be difficult to predict especially for infants born preterm. As ^{1}H NMR and UPLC-MS techniques are ideal for metabolic profiling of biofluids to gain complementary data from the same sample it was important to produce a protocol to extract faecal water for both analytical techniques using one faecal sample.

For UPLC-MS, the use of organic solvent is important in biofluid preparation to ensure the precipitation of proteins from the sample. From previous work into the optimisation of solvent extractions for faecal samples, H_2O was found to extract the highest volume of metabolites and produce the best resolution ^{1}H NMR spectra. Therefore, a sample that was extracted using ACN:H_2O was compared to the same sample extracted using H_2O only. From the results in Figure 3-20b & 3-20e, there was no differences in the raw ^{1}H NMR spectra in metabolite extraction or resolution.

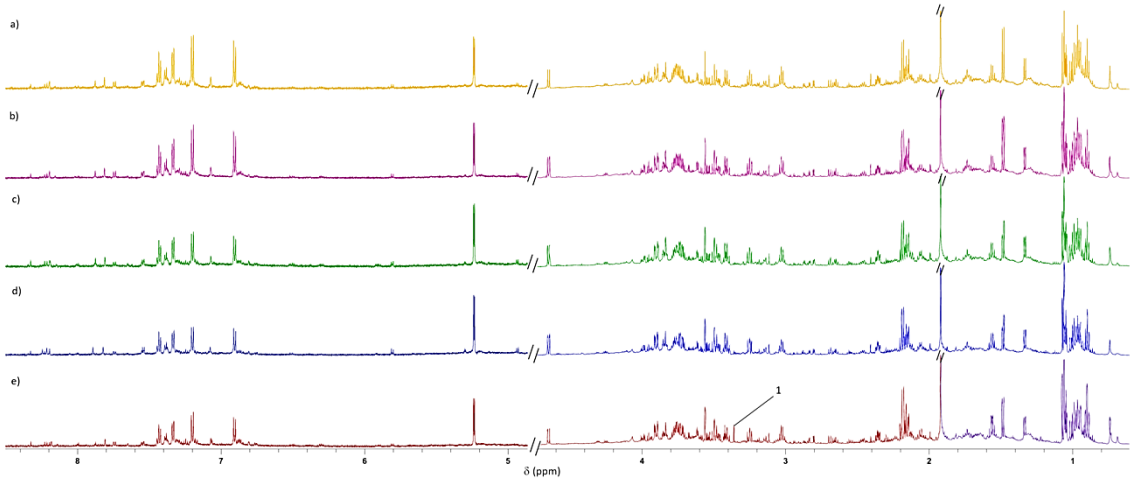


Figure 3-20: 1D ^{1}H NMR spectra of infant faecal water extractions using a) Acetonitrile:H_2O (1:3, ACN:H_2O), filtered and speed vacuumed; b) a) Acetonitrile:H_2O (1:3, ACN:H_2O), non-filtered and speed vacuumed; c) H_2O, filtered and speed vacuumed; d) H_2O, filtered; e) H_2O, non-filtered. Key: 1. Methanol, δ3.36

Also, as samples which contain organic solvent need to be evaporated and reconstituted before they can be analysed using NMR , the same aliquot extracted using H_2O was compared using speed vacumning compared to the same sample which hadn't been subject to speed vacumning. From the results shown in Figure 3-20c & 3-20d, there was no differences in the raw ^{1}H NMR spectra in metabolite extraction or resolution. After this investigation, it was decided that evaporation under nitrogen gas would be more appropriate for high throughput studies.

Furthermore, samples undergoing UPLC-MS analysis, need to be passed through a 0.2μm filter. Therefore, the same faecal sample from both solvent extraction techniques were subject to filtering and not filtering before analysis. From the results in Figure 3-20, the sample which had not been filtered or speed vacumned had retained the metabolite methanol (δ3.36). This metabolite is very volatile so this is not surprising but other volatile metabolites, including SCFAs weren't affected. Therefore, this method was accepted for ease of high through put and minimal sample volume.

3.7. DISCUSSION

This chapter has demonstrated the need for standardisation of biofluid collection and preparation in new-born and infant studies. Infant biofluids are a valuable resource to understand the development of the new-born infant after birth, therefore it is very important to prepare them correctly so that as much reliable information as possible can be obtained.

3.7.1. EFFECTS OF DILUTION FACTORS AND NORMALISATION IN INFANT BIOFLUIDS METABOLIC PROFILES

From these results, it is clear that metabolic information can be gathered from small amounts of sample volume for metabolic profiling using ^{1}H NMR spectroscopy and UPLC-MS. However, as the sample volume decreases the amount of specific biochemical information from the sample also decreases. Also, from the different biofluids assessed in this chapter, it is apparent that contamination is still a very clear problem for metabolic profiling and by introducing solvents to biofluids this can easily introduce complications to a dataset. A more accurate way of controlling differential concentrations between samples would be to ensure that all samples were adjusted for osmolality prior to analysis. However, this would introduce an extra time constraint and potential for human error, especially with larger projects.

Another factor to consider is that as the volume of sample decreases, the amount of FIDs acquired in each experiment increased, and therefore the length of the experiment increased, which reduces the efficiency of the metabolic screening process. This time was taken in consideration when the final protocol was produced, especially as longer experimental times are impractical for high throughput of samples and make huge demands on equipment resources.

From the results, it was clear that metabolic profiles from urine samples at half the recommended volume could be used in studies if necessary. With the increased number of scans and the use of normalisation techniques, samples give useful metabolic data.

For this investigation plasma samples from infants was unfortunately too difficult to obtain and therefore adult plasma was used. Due to this, it is difficult to conclude if infant plasma samples are appropriate for metabolic profiling. Further investigation into the use of infant dried blood spots for ^{1}H NMR analysis for metabolic profiling would be very advantageous. Furthermore, plasma samples in this investigation showed that water suppression issues and biological interactions attributed to protein binding make this biofluids more complicated when diluted.

Faecal samples act similarly to urine samples when a smaller mass is used. It is clear that half the recommended amount of faecal sample can be used (100mg), combined with a higher number of scans, when acquiring using ^{1}H NMR. Furthermore, from the results, good metabolite recovery is found in samples at a quarter the recommended amount of faeces for extraction also (50mg). Therefore, samples at very low volumes can be acquired using ^{1}H NMR if necessary; this is promising for infant studies where this would be a common issue.

3.7.2. AGE VARIABILITY AN IMPORTANT FACTOR WHEN ASSESSING INFANT METABOLIC PROFILES

Throughout pregnancy the developing foetus relies on the mother for nutrient uptake, waste removal and gas exchange and therefore will demonstrate strong similarities in the metabolic profile (Tea et al 2012; Lager et al 2012). Regularly timed urine samples from infants, taken at birth through the neonatal period and onto infancy, highlights the high variability in urine composition in relation to age.

Infant age is an important variable that contributes to the metabolic profile of biofluids and thus establishing the metabolic profiles linked to age needs to be a priority for future metabonomic studies. From birth, the infant metabolic profile can be related to their mothers and goes on to change dramatically after birth due to environmental factors. As the infant's body develops and their gut microbiome quickly establishes, the metabolic profile reiterates this through changes in their metabolome. The knowledge of an age-related metabolic profile can make the interpretation

of data for other effects such as disease easier, which is important in biomarker discovery for disease diagnosis or prognosis. It shows that importance of obtaining samples from infants at the same time point, as even a few days or a week can completely change the metabolic profile and cause confounding results. In studies where age cannot be standardised, this should be appreciated and incorporated into the data modelling strategy.

3.7.3. ASSESSMENT AND CHARACTERISATION OF CONTAMINATION FROM COLLECTION MATERIALS IN INFANT URINE

Clean catch urine is difficult to obtain from infants and thus typical collection means include placing adhesive sample collection bags on the infants or more practically relying on extracting urine from cotton wool balls placed inside the nappy. Both cotton wool and nappies contain chemical contaminants, which contribute to the urine metabolic profile and which can easily be detected by both ^{1}H NMR and UPLC-MS techniques. It is necessary to recognise these contaminants before studying urine data so that their signals can be excluded from the statistical models.

This issue of possible contaminants has been examined and further reported using complementary ^{1}H NMR and UPLC-MS-based metabonomics of three nappy brands using the 'nappy plus cotton ball' technique. It was found that the NMR spectra of the different cotton ball brands examined were identical in composition, thus only one brand of cotton ball was used for 'nappy plus cotton ball' samples analysis. It also indicated, that certain nappy brands are more likely to exhibit distinct contaminant profiles than others; for example Pampers contains acrylic acid. Therefore, it is advised that nappies and cotton balls are characterised using metabonomic methodologies prior to urine samples analysis. This way contaminants coming from any material would be manageable and unwanted variation introduced by contamination could be minimised.

Although some of these contaminants have been previously reported, such as polyethylene glycol (PEG) from nappies and propylene glycol in infant urine which has been suggested to derive from babywipes, this is a more comprehensive assignment of exogenous chemicals deriving from

collection procedures (Goodpaster et al 2011; O'Sullivan et al 2013). A systematic analysis of the effect of different nappies on synthetic urine by Goodpaster and colleagues (2011) showed that multivariate analysis of either NMR or LC-MS spectra resulted in clustering of spectra according to the brand of nappy. In Goodpaster's study, it was estimated that up to 59% of the NMR spectrum was influenced by chemical contamination from diapers. Thus, it is important that the brand of diaper is standardised across a study to prevent the introduction of extraneous variation.

This chapter has demonstrated the importance of assessing materials associated with the collection and preparation of infant biofluids. It needs to be ensured that introduced contaminants are well characterised so that discriminant metabolites which are products of neonate's metabolism, kidney function or gut microbiome co-metabolites can be recognised.

3.7.4. OPTIMISATION OF INFANT FAECAL WATER EXTRACTION FOR METABOLIC PROFILING

NMR-based faecal water metabolic profiling offers a great opportunity for understanding the complex interaction between humans, their diets and their gut microbiome (Saric et al 2008; Nicholson et al 2012). Furthermore, the extraction of faecal water for metabolic profiling has not previously been optimised for infant biofluids where volume of sample is limited. In this chapter, the optimisation for faecal water to be acquired by ^{1}H NMR was produced to measure a broad range of different metabolites and obtain a global overview on diet and health of infants as well as interaction with the gut microbiome. It was also important to produce a protocol for high throughput and large dataset studies.

To ensure a high recovery of metabolic information using NMR based metabolic profiling, stability of the sample and reproducibility of information is maintained, the selected method involves a number of different steps. Before homogenisation, it is recommended that faecal samples are subject to two freeze thaw cycle, this can release higher concentrations of all the metabolites, but more than two freeze thaw cycle can start to cause degradation of the metabolic profile.

To extract the water from the faecal sample, bead beating using zirconium beads should be used for the best results and for high throughput samples. For smaller sample sets, homogenising the samples using pestle and mortar whilst frozen under liquid nitrogen produces the best result for concentration of metabolites.

For extraction, solvents should be added to the faecal samples before homogenisation to capture metabolites in the liquid for later analyses. In this investigation, a number of different solvents were tested and water is believed to be the best solvent for extraction. When using deuterated solvents, including phosphate buffer, metabolite concentration is considerably lower and loss of information is possible. This may be due to some macromolecules which don't dissolve properly in the deuterated solutions bind to smaller molecules and therefore less information is available for metabolic profiling.

To ensure that infant faecal samples could undergo more than one analytical techniques, extraction of these samples were also tested for the possibility to extract for ^{1}H NMR and UPLC-MS. By using an organic solvent and filtering when extracting, this was possible. Adding these elements to the protocol did not introduce any contaminants or affect the metabolic profile except for the loss of methanol. With this in consideration, the ease of the protocol for extraction for both platforms and the ability to extract for high through put studies meant that this was acceptable.

3.8. CONCLUSION

Currently most clinical tests available in neonatology rely on simple technologies that measure chemicals in blood, urine or other biofluids; including blood glucose, acid-base profile, lactate and ammonia. These tests can help to identify known groups of metabolic disorders; however, these tests are neither sensitive nor specific for any particular disease. The goal is for each condition or disease to acquire a discriminatory metabolic phenotype, derived essentially from small but significant differences in certain metabolites. Thus, scientists need to be sure that these

discriminant metabolites are products of a neonate's metabolism and kidney function and not just contaminants of the cotton wool and nappies used to collect the biofluids.

A reproducible and well established metabolic profiling protocol for infant biofluid sample collection and preparation needs to be introduced. The analysis of this data as well as integration of this data with other 'omic' analysis will make this technology valuable in the clinical setting for disease diagnosis. Metabonomic research in neonatology offers a great potential to improve the diagnosis and treatment of infant's diseases, significantly decreasing the stress that newborns endure. Metabolic profiling of infants may deliver validated biomarkers which can be used for predicting individual predispositions and disease etiology or to the monitoring of disease's progression and treatment management.

CHAPTER 4

NORWEGIAN MICROBIAL STUDY (NoMic): A TIME SERIES INVESTIGATION OF INFANT DEVELOPMENT

4.1. INTRODUCTION

We provide the bacteria with a stable environment, and in return the bacterium produce metabolites and vitamins from our diet. The gut is quickly colonised after birth, and this microbial community develops through the first few months of life (Nicholson et al., 2012). It is believed that gestational age, the mode of delivery, nutrition (e.g. maternal or formula milk), clinical interventions (e.g antibiotics), environmental exposures as well as maternal microbiota can affect the composition of the gut microbiome (Mueller et al., 2015).

Faecal metabolic profiling has become a useful tool for deepening our knowledge into the interactions between humans, nutrition and microbiome communities. A number of different studies have utilised faecal samples in adults to predict risk for diseases including inflammatory bowel disease and colon cancer (Marchesi et al., 2007; Weir et al., 2013; Le Gall et al., 2011). There are few studies on infant faecal metabolic profiling but recent reports have shown promise in using

this biofluid as a tool for further understanding interactions among diet, microbes, and host (Chow et al., 2014; Martin et al., 2014).

The gut microbiota is made up of a complex society of microbes which impacts the body by nutrient processing, vitamin synthesis and playing a key role in the maturation of the immune system. In recent years, obesity at childhood has been on the rise (Biro et al., 2010) and the gut microbiota has been indicated in the development of obesity (Koleva et al., 2015). It has been observed that childhood obesity is an important predictor of adult obesity, morbidity and mortality and with the number of people developing obesity increasing exponentially worldwide it is very important to understand the early development of the disease.

Studies on germ-free mice have provided evidence for the association of body weight and gut microbiota; including evidence for the existence of an 'obese microbiome' that has the ability to harvest increased amounts of energy from the diet (Ley et al., 2005, Ley et al., 2006). Obesity alters the gut microbiome by reducing the diversity of the community as well as altering metabolic pathway and gene expressions (Turnbaugh et al., 2006). This development is thought to be influenced by early life exposures, including nutrition and early rapid weight gain, that alters the composition of gut microbiota (Arenz et al., 2004, Koletzko et al., 2012). A study using metabolic profiling of children with obesity discovered that BCAA and hormones were associated with adiposity and cardiovascular risk during childhood (Perng et al., 2014); they went on to discover that maternal BMI status may also contribute to this profile.

Metabolites are representative of the underlying biological development of obesity and therefore may have a major impact on outcome. By understanding the early development of the metabolic profile it will build up towards understanding early risk factors for obesity later in life.

4.2. AIM

It is has become increasingly apparent of the importance to understanding metabolite composition in infancy and how influential the gut microbiota is during this developmental period. This study proposes to investigate the early development of the infant metabolism and the influence of the gut microbiome.

To do this the Norwegian Microbial (NoMic) study focussed on infants from birth up until the age of two (Eggesbo et al., 2011). The NoMic study was funded by a grant from the South-Eastern Norwat Regional Health Authority and in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. Ethics for this project was approved by the Regional Committee for Medical Research Ethics, south East Norway (2012/1259-19). The work conducted in this thesis was funded as a pilot study and therefore the costs incurred were covered by the author's department. My role was to perform the metabolic phenotyping of the NoMic stool samples from the infants using ^{1}H NMR; this work had two main objectives:

1. Characterise the faecal metabolic profile's time trajectory in infants to establish a foundation for future investigations into the early metabolome.

2. Understand the influence of early life experiences on the metabolic profile to identify early risk factors and potential early biomarkers for the prevention of metabolic disorders including obesity in later life.

For this project, the characterisation of the developmental trajectory of the faecal metabolic profile is key to understand the baseline metabolism in the early years of life. Therefore, time series analysis was necessary to examine this data as there are multiple samples taken from the same individual through time as well as a large cohort of samples. This analysis will be used to identify metabolic features significantly altered through time between different groups, including term versus preterm at birth, mode of delivery, and BMI status at 1 and 2 years of age.

4.3. MATERIALS AND METHOD

4.3.1. NORWEGIAN MICROBIOME (NOMIC) STUDY SAMPLE SET INFORMATION

Faecal samples were collected at six time points; 4 days, 10 days, 30 days, 120 days, 1 year, and 2 years after birth; comprising of 1802 faecal samples used for this thesis. Five hundred and twenty-four healthy infants were recruited to take part in this non-intervention clinical study. A faecal sample was taken on at least one time-point from each child, if not more, but not all individuals have provided a sample at all six time points. A detailed description of how many samples were collected from each time point can be found in table 4-1 as well as further clinical information infants and mothers.

Relevant clinical data were collected for each infant at each time point and questionnaires on nutritional information were also obtained; including months that each child was exclusively fed breast milk. This included information on gestational age and weight at birth of the infant; including mode of delivery of the infant. Information on allergies, asthma and colic. Further to this, antibiotic treatments prescribed to the child and at year 1 and 2, the infant's BMI status was recorded.

Maternal information was also noted including smoking before or after the pregnancy, any antibiotic treatments during the pregnancy and during or after birth. Maternal age, ethnicity as well as BMI were also recorded at the time of the birth.

Table 4-1: Clinical data information for Norwegian Microbial (NoMic) sample set at each time point and all the time points together. Data is expressed as the median (IQR) or as a percentage of subjects.

Clinical data	Sample time point						
	All samples	**4**	**10**	**30**	**120**	**365**	**730**
Number of samples	1802	287	291	272	239	405	308
Number of infants	512						
Infants with samples at all 6 timepoints	56 (11.6%)						
Birth gestational age (range)	39.7 (23.3-44.5)	39.85 (27.3-44.5)	39.71 (27.3-44.6)	39.64 (27.3-44.6)	40.1 (27.5-43.9)	39.57 (23.3-44.5)	39.57 (25.3-44.6)
Preterm samples (%)	26.47%	23.34%	28.50%	28.40%	17.15%	27.65%	30.50%
SGA (%)	13.28%	12.50%	12.70%	14%	10.80%	13.30%	15,6%
Weight median (g)	3370 (500-5050)	3400 (832-5050)	3365 (832-5050)	3352 (832-5050)	3540 (920-5050)	3370 (500-4695)	3290 (652-5050)
Caesarean section (%)	31.96%	29.26%	33.30%	33.90%	33.47%	31.85%	29.80%
Breast feeding exclusively (months)	4 (0-12)	4 (0-11)	4 (0-11)	4 (0-12)	4 (0-12)	4 (0-12)	4 (0-11)
Gender (% male)	45.90%	45.64%	45.70%	43.75%	46.86%	50.10%	46.75%
BMI 1 yr (% high)	19.72%	19.32%	20.37%	20.50%	20.60%	19.20%	19.65%
BMI 2 yr (% high)	18.47%	20.15%	15.82%	16%	19%	18.26%	21.80%
Allergies (%)	10.45%	8.70%	9.62%	8.80%	9.60%	12.10%	12.66%
Asthma (%)	8.33%	8.01%	8.90%	9.20%	9.20%	6.66%	8.76%
Colic (%)	14.65%	15.13%	13.85%	17.64%	14.07%	13.10%	15.60%
Infant antibiotic prescribed in 1st 2 years	61.7%%						
Maternal age (yrs)	30 (20-40)						
Maternal pre pregnancy BMI	24.09 (16.45-47.89)						
Maternal post pregnancy BMI	29.4 (17-51.5)						
Maternal ethnicity (Norwegian)	88.42%						
Maternal smoking pre pregnancy	11.9%%						
Maternal smoking post pregnancy	8%						
Maternal antibiotic taken during pregancy	0.80%						
Maternal antibiotic taken during birth	6%						
Maternal antibiotic taken after birth	14.50%						

4.3.2. FAECAL WATER PREPARATION

Faecal samples were weighed out (100mg) and transferred to microvials which contained 0.1mm zirconium beads whilst frozen. They were then subject to two freeze thaw cycles before having 250ul of ACN:H_2O added to the microvials. These vials were homogenised using a bead beater for 10 seconds and centrifuged at 16,000 xg for 10 minutes, at 4°C. The supernatant was taken from the vial and placed in pre-washed spin columns and again centrifuged for 20 minutes at 16,000 xg, at 4°C. The faecal water was evaporated under nitrogen gas for 30 minutes to remove any organic solvent. Once evaporated, these samples were reconstituted into 540µl D_2O and 60µl phosphate buffer, vortexed and sonicated for 5 minutes. The faecal water was then transferred into 5mm NMR tubes ready to be acquired using a 600MHz ^{1}H NMR spectroscopy.

4.3.3. ^{1}H NMR SPECTROSCOPY OF FAECAL WATER

^{1}H NMR spectra were acquired on a Bruker DRX-600 spectrometer (Bruker Biospin, Karlsruhe, Germany) operating at 600.29MHz for proton observation using a standard one dimensional water pre-saturation pulse sequence [relaxation delay-90°-t1-90°-tm-90°-acquire free induction decay (FID)]. The relaxation delay was 4 seconds; with application of a 90° radio frequency pulse, t1, referring to the interpulse delay, which was set to 3µl, while tm is the mixing time of 100ms. The probe was matched and tuned automatically to the proton transmitter resonance frequency before acquisition for each sample and samples were run at a temperature of 300K.

Processing of ^{1}H NMR spectra was carried out using TOPSPIN 3.1 software package (Bruker Biospin, Rheinstetten, Germany). The FIDs were transformed into spectrum by Fourier transformation. The spectra were manually phased, baseline corrected and calibrated to the TSP signal at δ0.0 for faecal samples.

4.3.4. MULTIVARIATE DATA ANALYSIS

4.3.4.1. ^{1}H NMR data pre-processing

The spectral data (from δ0.5 to 9.5) were imported into Matlab software (version 2014a, the Mathworks Inc, MA, USA). The spectral region incorporating the water (δ4.6-5.05) was also removed from each spectrum. ^{1}H NMR spectra were aligned using an in-house algorithm and normalised using a Probalistic Quotient normalisation (PQN) in order to remove variation in metabolite linked to osmolarity and dilution.

4.3.4.2. ^{1}H NMR metabolic data analysis

Multivariate statistical models (PCA, OPLS, OPLS-DA) were performed in Matlab software (version R2014b, the mathworks Inc, MA USA) and SIMCA-P 14 software (Umetrics, Umea, Sweden) using univariate scaling. The $Q^{2}Y$ value was determined for each model by plotting the data against the cross validated scores to estimate the goodness of fit and level of over fitting for the OPLS and OPLS-DA models. All models were subjected to a 7-fold cross validation and were validated by multiple permutation tests (permutation number =1000). (see methods Chapter)

4.3.4.3. A Time Series Analysis of ^{1}H NMR faecal metabolic data

^{1}H NMR spectral faecal data were binned into 1000 bins producing variables, which represented 0.01ppm scale rather than 0.001ppm. An in-house time-series algorithm that relies on Functional Spline for Short Asynchronous Time-Series Analysis was used to identify faecal metabolic features that are significantly altered in time between groups (Ramsay et al., 1988). These groups included mode of delivery, term versus preterm and BMI status of infants at 1 and 2 years of age. Significant differences in variables between time points from the groups were identified using p-values of less than 0.01 and trajectory of the variables were observed using spline graphs. Further explanation on this method can be found in the methods chapter section.

4.4. RESULTS

4.4.1. ^{1}H NMR spectroscopy metabolic profiling

Information was collected from both infant and mother through questionnaires and recording medical records during the sampling of infants. Statistical analysis was performed on faecal samples as a whole data set as well as individual time points to accommodate the complexity that time adds to the data given that development of the microbiota is an ongoing process that doesn't necessarily co-evolve at the same rate in all children. Characterisation of the development of the infant faecal metabolic profile is the main aim of this project. Further to this, I investigated the effect of preterm birth, the mode of delivery, BMI status, as well as the time of exclusivity of breast feeding has on the faecal metabolic profile through this highly developmental stage.

4.4.1.1. ^{1}H NMR faecal metabolic profiling of developing infants from 4 days until 2 years postpartum.

Using the faecal ^{1}H NMR metabolic profiles, a principal component analysis (PCA) was performed (Figure 4-1), which demonstrated a clear separation of samples according to their assigned class; six time points from 10 days to 730 days postpartum. This model was comprised of all the time points and was formed of three principal components, which collectively explained 34% of the variability within the metabolic data. The optimal number of orthogonal components was determined by R^2X (goodness of fit) and Q^2Y (predictability of fit) statistics.

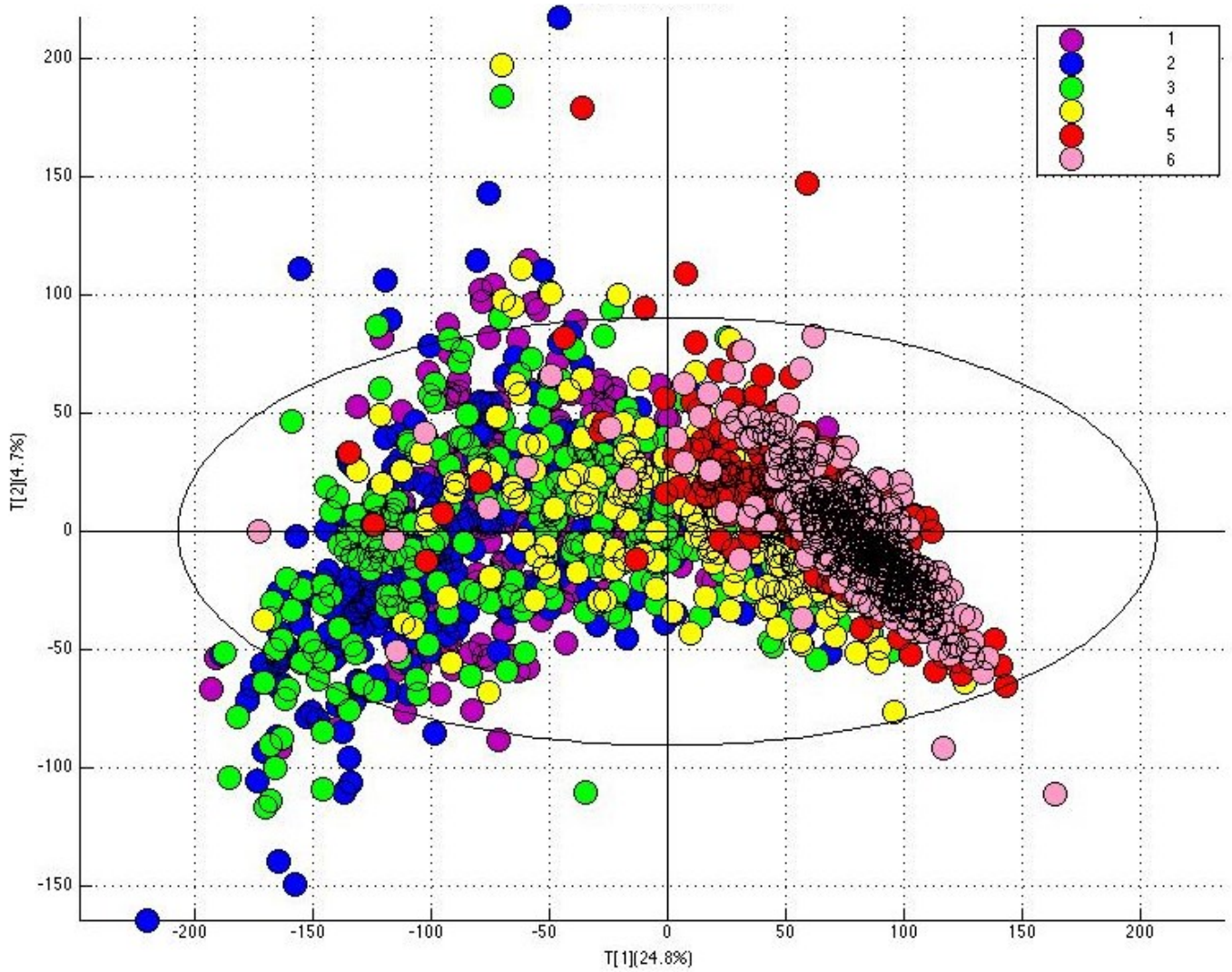


Figure 4-1: PCA of 1D ^{1}H NMR faecal spectral data from all infants at all time points. Key: Purple = Day 4, Blue = Day 10, Green = Day 30, Yellow = Day 120, Red = Day 365, and Pink = Day 730; Component 1 explains 24.8% of variance whereas component 2 explains 5%. R^2X (cumulative) = 0.34 and Q^2X (cum)=0.33.

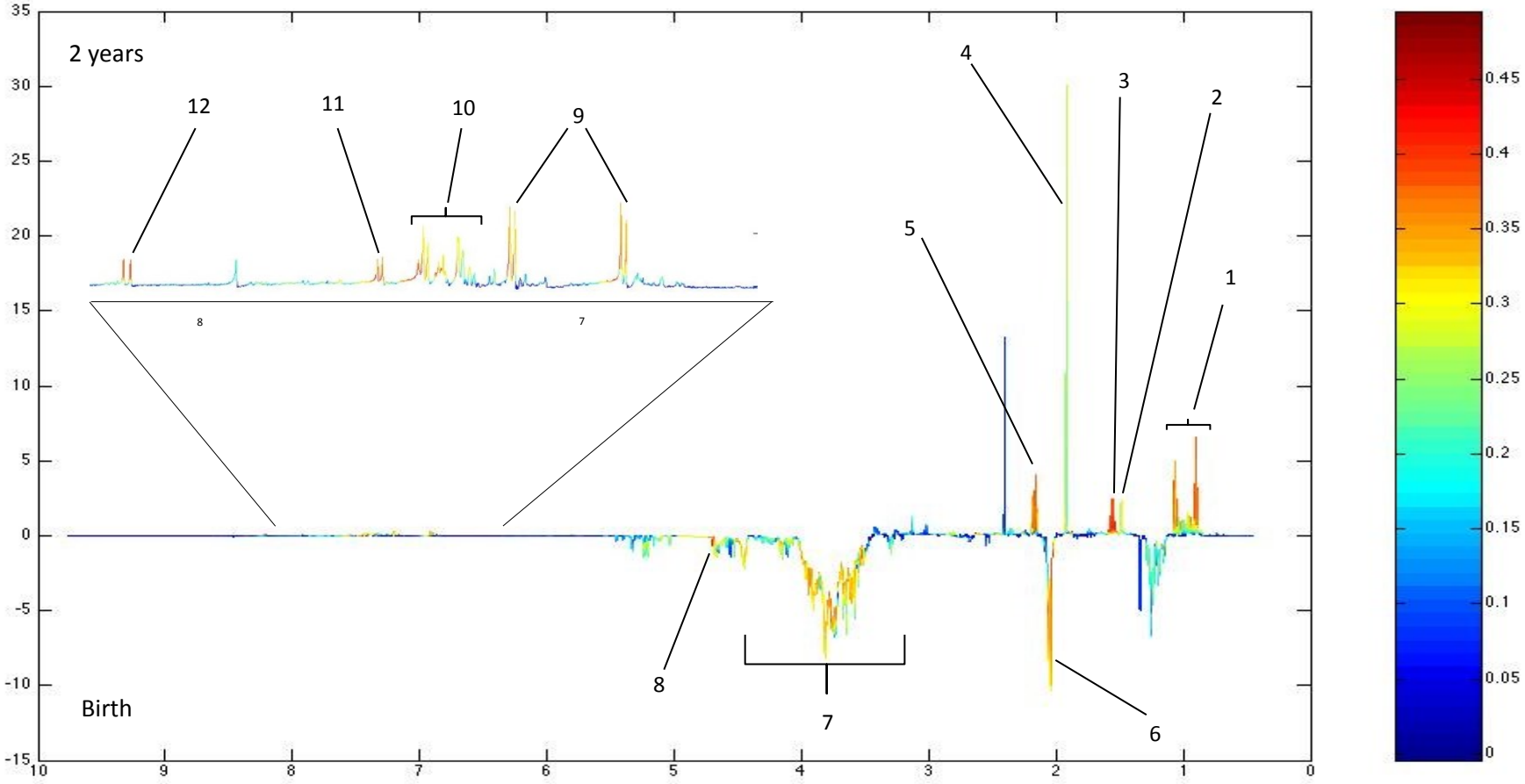


Figure 4-2: Correlation Coefficient plot from OPLS regression model (R^2X= 0.32, Q^2X=0.73, R^2Y = 0.78) showing significant metabolites in faecal water separating between infants at different time points: 4, 10, 30, 120, 365 and 730 days postpartum. 1. Butyrate (δ 0.89), Propionate (δ 1.06), & Valine (δ1.02), 2. Alanine (δ1.48), 3. Butyrate (δ1.56), 4. Acetate (δ1.92), 5. Butyrate (δ2.17) & Propionate (δ2.18), 6. *N*-acetyl-glycoprotein fragments (δ2.02, δ2.06), 7. Sugars (δ3.54, δ3.62, δ3.75, δ3.87, δ3.94, δ4.0), 8. Glucose (δ4.68), 9. Tyrosine (δ6.91, δ7.2), 10. Phenylalanine (δ7.38, δ7.43), 11. Uracil (δ7.55), 12. Adenine (δ8.2, δ8.23)

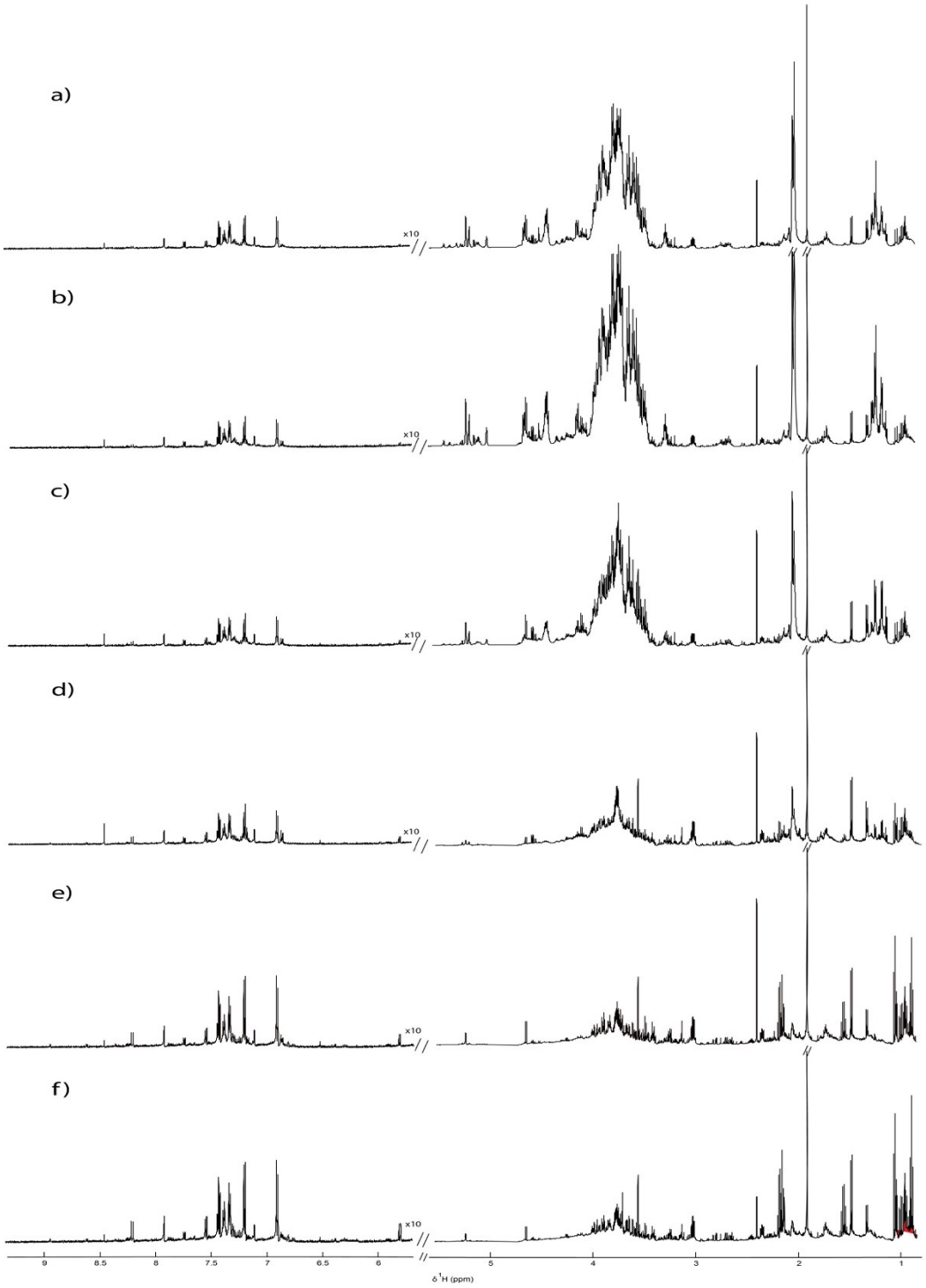


Figure 4-3: Median 1D ^{1}H NMR faecal spectral data from all infants at all time points. A) Day 4, b) Day 10, c) Day 30, d) Day 120, e) Day 365, and f)Day 730; (Aromatic region (5.5-9.5ppm) of the spectra has been increase by x10)

However, from Figure 4-1, day 4 postpartum faecal samples aren't as well separated within the data and are similar to samples taken at 30 days postpartum. Considering the median spectra in Figure 3, day 4 postpartum samples seem to have relatively lower concentration of metabolites compared to day 10, especially the aliphatic region of the spectrum where carbohydrates and sugar are found as well as *N*-acetylglycoproteins fragments found around δ2.04 (data not shown). This higher concentration of metabolites could be the influx of enteral feeding in the first few days not being recognisable in the metabolic profile until after the first week of life. From the PCA scores plot (Figure 4-1), the most noticeable difference between the time points is the variation in the early samples is much higher compared to samples taken at 1 and 2 years of age.

An Orthogonal Partial Least Squares (OPLS) regression model was performed on all faecal data using time points as a regression class. For this model the metabolites with the highest correlation coefficient were identified (Figure 4-2) and metabolite assignments are summarised in table 4-2. In particular, a metabolite profile high in carbohydrates, sugars (δ3.54, δ3.62, δ3.75, δ3,87, δ3.94, δ4.0, δ4.68) and *N*-acetylglycoprotein fragments (δ2.02, δ2.06, δ4.45), including *N*-acetyl glutamate (δ4.16) can be observed in samples obtained at earlier time points, (e.g. days 4 and 10). Whereas, faecal water acquired at later time points (e.g year 1 and 2) are more strongly associated with organic acids including, amino acids (alanine (δ1.48), tyrosine (δ6.91, δ7.2) and aspartate (δ2.79, δ2.83), phenylalanine (δ7.38, δ7.43)), SCFAs (butyrate (δ0.89, δ1.56, δ2.17), propionate (δ1.06, δ2.18) and acetate (δ1.92)), and BCAAs (valine δ1.02). Later samples are also associated with succinate (δ2.4), methylamine (δ2.6), uracil (δ5.8, δ7.55) and adenine (δ8.2, δ8.22).

Table 4-2 – List of discriminatory metabolites assigned from OPLS time point regression model (Figure 4-2). (*overlapped resonances from glucose, other sugars and polyols, *N*-methylated compounds and amino acids)		
Metabolite name	**^{1}H NMR δ (multiplicity group)**	**Days postpartum**
N-acetyl glycoprotein fragments	2.02 (s), 2.06 (s)	≤30 days
Sugars*	3.54 (m), 3.62 (m), 3.75 (m), 3.87 (m), 3.94 (m), 4.0 (m)	≤30 days
N-acetylglutamate	4.15 (m)	≤30 days
Glucose	4.68 (d)	≤30 days
Butyrate	0.899 (t), 1.56 (m), 2.17 (t)	≥120 days
Propionate	1.06 (t), 2.18 (q)	≥120 days
Valine	1.02 (d)	≥120 days
Alanine	1.48 (d)	≥120 days
Unknown 1	1.87 (s)	≥120 days
Acetate	1.92 (s)	≥120 days
Succinate	2.4 (s)	≥120 days
Methylamine	2.60 (s)	≥120 days
Aspartate	2.83 (dd)	≥120 days
Unknown 2	2.63 (s)	≥120 days
Uracil	5.8 (d), 7.55 (d)	≥120 days
Tyrosine	6.91 (d), 7.2 (d)	≥120 days
Phenylalanine	7.38 (m) 7.43 (m)	≥120 days
Adenine	8.2 (s) 8.23 (s)	≥120 days

The median spectrum from each time point reiterates this variability in faecal sample metabolic profiles clearly (Figure 4-3), with a dramatic change between faecal samples taken at 30 days postpartum compared to 120 days postpartum. The highest correlating metabolites in infant faecal samples at day 30 compared to day 120 using an OPLS-DA model are detailed in table 4-3. Faecal samples taken at 30 days are highly associated to metabolites found in the carbohydrate region of the NMR spectrum (δ3.51, δ3.61, δ3.87, δ4.14, δ5.11, δ5.24) as well as *N*-acetyl glycoprotein fragments (δ2.02, δ2.06, δ4.45) and faecal samples taken at 120 days postpartum, are highly associated with bile acids (δ0.68), methylamine (δ2.6) and 4-cresyl glucuronide (δ2.32) as well as an unidentified metabolite (δ1.65, δ2.24).

Table 4-3 – List of discriminatory metabolites assigned in the loadings plot of OPLS-DA model of infants at time points 30 day versus 120 day postpartum (*overlapped resonances from glucose, other sugars and polyols, *N*-methylated compounds and amino acids)		
Metabolite name	**1H NMR δ (multiplicity group)**	**Time point**
N-acetyls of glycoprotein fragments	2.02 (s), 2.06 (s)	Day 30
Sugar*	3.51 (m), 3.61 (m), 3.87 (m), 4.14 (m), 5.11 (d), 5.24 (d)	Day 30
Bile acids	0.68 (s)	Day 120
Unknown 1	1.65 (m), 2.24 (dd)	Day 120
4-cresyl glucuronide	2.32 (s)	Day 120
Methylamine	2.6 (s)	Day 120

4.4.1.2. ^{1}H NMR faecal metabolic profiling of infants highlights differences between infants born term vs preterm

Preterm infants are classed as born at less than 37 weeks gestation; however, there are different severities of prematurity depending on the gestational age (GA) at birth. Further categorising of preterm infants into their severity was established for this model; early preterm (GA ≤31 weeks, n=121), late preterm (32-36 weeks, n=355), and term (>37 weeks GA, n=1325). An OPLS model, using one orthogonal component, was produced from the whole dataset to assess the overall differences between the term and different severities of preterm infant samples. The discriminatory metabolites associated with term versus preterm born infant samples are summarised in table 4-4. The final R^2Y and Q^2Y values equalled 15% and 9% respectively. The Q^2 value for this model is fairly low, which is due to the nature of the high variance of stool metabolic profiles.

There is a positive correlation to butyrate (δ0.87, δ1.59) in the early and late preterm born infant samples compared to the term born babies. These samples also related to an unknown metabolite (δ1.38). and a metabolite found at δ1.31 which has been tentatively assigned as the signal coming from the fatty acid chain of a triacylglyceride and will be named lipid 1 (L1) for the remainder of this thesis.

Comparatively, term infants had relatively higher concentrations of sugars (δ3.46) and specifically glucose (δ4.59 and δ5.27) compared to preterm samples. Furthermore, infants born at term were highly associated with choline (δ3.28, δ3.53, δ4.07) and *N,N,N*-trimethyllysine (δ3.207).

Differences in metabolite composition between infants born term and early or late preterm was similar when the data was modelled separately for each time point compared to modelling the whole data set. This would suggest that these samples gave the largest variation in the data and may have been the drivers for the models when all samples were included. When separating the data set into specific time-points and modelling the data using OPLS, there were clear differences in faecal metabolic profile from day 4 postpartum through to 4 months of age; discriminatory metabolites are summarised in table 4-6. However, faecal samples taken at days 365 and 730 did not show any significant separation in samples taken from term versus preterm infants indicating that early life differences are reflected in metabolite profiles but over time these differences diminish.

Faecal samples taken from early and late preterm infants at day 4 were highly associated with 3'-fucosyllactose (δ1.18, δ5.39, δ5.47), which is a common breast milk oligosaccharide (Practico et al., 2013). Whereas, infants born at term were more associated with various bile acids (δ0.728, δ0.735) and a sugar found at δ3.46.

The meconium is the first stool of the infant and is the contents of the foetal gastrointestinal tract. At birth the intestine sheds its protective layer consisting of bile, epithelial cells and hair which makeup the meconium. However, the meconium is usually excreted within 24 hours in full term infants (Clarke, 1977), therefore, these samples taken at 4 days postpartum are not meconium but suggest that there are still elevated secretion concentrations of bile acid in the first few days in term compared to preterm infants.

Discriminatory metabolites associated with the two classifying groups samples change by the time the infant is 10 days postpartum. Infants born preterm were highly associated with L1 (δ1.31) and

tyramine (δ7.23) in their faecal metabolic profile. However, term born infants were correlated with choline (δ3.28, 3.53, δ4.07), glucose (δ4.59, δ5.27) and *N,N,N*-trimethyl-lysine (δ3.207).

At 30 days' postpartum, infants born preterm had metabolic profiles correlated to butyrate (δ0.86, δ1.59) and two unidentified metabolites (δ1.38, δ1.99). Infants born term, when compared at this time point, separated from preterm infants due to choline (δ3.28, 3.53, δ4.07) and tyrosine (δ6.91, δ7.2).

At 120 day's postpartum, the faecal metabolic profiles from early and late preterm infants separated due to the higher association with sugars (δ3.85, δ4.01). Term samples were correlated to a number of different metabolites including leucine (δ0.96), alanine (δ1.48) and succinate (δ2.41). Furthermore, term infant's faecal metabolic profiles was highly associated to choline (δ3.28, 3.53, δ4.07), tyramine (δ7.23) and phenylalanine (δ7.43).

Table 4-4: List of discriminatory metabolites assigned from multivariate models of infant ^{1}H NMR spectral data; Early Preterm (<27wks GA), Late Preterm (23-36wks GA) and Term (>37wks GA) Infants, Birth gestational age (GA) & Birth weight (OPLS-regression) – High (H) vs Low (L) (*overlapped resonances from glucose, other sugars and polyols, *N*-methylated compounds and amino acids)

Metabolite Name	^{1}H NMR δ (multiplicity)	Early Preterm, Late Preterm & Term Infant					Birth GA - High vs Low					Birth Weight - High vs Low				
		Time point					Time point					Time point				
		All	4	10	30	120	All	4	10	30	120	All	4	10	30	120
Bile acids	0.728 (s), 0.735 (s)		T										H	H		
Butyrate	0.866 (t), 1.59 (m)	P			P		L			L					L	
Leucine	0.96 (t)					T					H					
Valine	0.99 (d)										H					H
Propionate	1.07 (t)								L	L				L	L	
Lipid 1 (L1)	1.31 (bs)	P		P			L		L	L		L		L	L	
3'-fucosyllactose	1.18 (m), 5.39 (d), 5.47 (d)		P					L	L				L	L	L	
Unknown 1	1.38 (dd)	P			P		L			L		H			L	
Alanine	1.48 (d)					T					H					H
Unknown 2	1.99 (s)				P					L						
Succinate	2.41 (s)					T					H					H
Unknown 3	2.967 (d)							L					L			
NNN-trimethyl-lysine	3.207 (s)	T		T			H		H	H		H	H	H	H	
Choline	3.28 (s), 3.53 (m), 4.07 (t)	T		T	T	T	H		H	H		H		H	H	
Sugar*	3.46 (d)	T	T				H	H	H	H			H			
	3.85 (s)					P					L					L
	4.01 (m)					P					L					L
Ethylene glycol	3.71 (s)						L	L	L			L		L	L	
Glucose	4.59 (d), 5.27 (d)	T		T			H		H			H				
Tyrosine	6.91 (d) 7.2 (d)				T	T				H	H				H	H
Tyramine	7.23 (d)			P					L						L	
Phenylalanine	7.43 (t)					T					H					

4.4.1.3. Faecal metabolic profiles from infants with variable birth gestational ages and birth weights

The OPLS model was calculated using four orthogonal components, which can indicate high variability between the samples; this is due to the orthogonal components required to remove the systematic differences caused by the natural variability, and not related to the class of interest. The final R^2Y and Q^2Y values equalled 44% and 22% respectively for birth gestational age. The final R^2Y and Q^2Y values equalled 34% and 13% respectively for birth weight.

The discriminatory metabolites associated with the two classifying groups for both the both birth gestational age and birth weight, models are summarised in table 4-4. Infants born with low birth weight and at a lower gestational age were highly associated with L1 (δ1.31), butyrate (δ0.86, δ1.59) and ethylene glycol (δ3.71). Infants born with a higher birth weight and a higher gestational age, their faecal profiles were positively correlated to choline (δ3.28, 3.53, δ4.07), *N,N,N*-trimethyl-lysine (δ3.207) and sugars (δ3.46), specifically glucose (δ4.59, δ5.27).

Interestingly, an unidentified metabolite - with a resonance of δ1.38 - was seen to be correlated to infants whom were born at a higher birth weight but was also highly associated with infants born at a lower GA. Unfortunately, without a more accurate identification of this metabolite a reasonable explanation for this difference cannot be concluded.

Samples taken at 4 and 10 days with a lower GA and birth weight were highly associated with ethylene glycol (δ3.71) and 3-fucosyllactose (δ1.18, δ5.39, δ5.47); whereas at 30 days postpartum this metabolite was only associated with a lower birth weight. At 4 days postpartum only, infants with a lower GA and birth weight are associated with an unknown metabolite (δ1.38). Samples from lower GA and birth weight were associated with propionate (δ1.07) and L1 (δ1.31).

Conversely, infants born at a higher gestational age, were more associated with sugars (δ3.85, δ4.01); but at a higher birth weight, these infants metabolic profile correlated with choline (δ3.28, 3.53, δ4.07) and *N*,*N*,*N*-trimethyl-lysine (δ3.207).

However, at 120 days' postpartum, the panel of faecal metabolic profiles separating the two classes was different. In lower birth gestational age and weight, infants were differentiated mainly by a higher correlation to sugars (δ3.46, δ3.85). In samples taken from infants at born at higher gestational age and birth weight, metabolic profiles were highly associated with the amino acids; alanine (δ1.48), valine (δ0.99) and leucine (δ0.96), as well as succinate (δ2.41) and tyrosine (δ6.91, δ7.2). Furthermore, infants that were born at later gestation age were more associated to phenylalanine (δ7.43).

4.4.1.4. ^{1}H NMR faecal metabolic profiling of infants highlights differences depend on the mode of delivery

A OPLS-DA model was produced from the whole dataset to assess the overall differences between the infants born vaginally compared to caesarean section infant samples; the discriminatory faecal metabolites associated with the two classifying groups were identified and are summarised in table 4-5. The O-PLS-DA model was build using 2 orthogonal components with the final R^2Y and Q^2Y values being 15% and 5% respectively.

From modelling all the samples, vaginally delivered infants were highly correlated with sugars (δ3.25, δ3.43, δ3.46) as well as alanine (δ1.48) and tyrosine (δ7.2). Acetone (δ2.22) was also present in relatively higher concentrations in faecal water taken from infants which were delivered vaginally.

Table 4-5– List of discriminatory metabolites assigned in the loadings plot of OPLS-DA model of infants delivered via different modes, vaginally (VD) and caesarean section (CS). (*overlapped resonances from glucose, other sugars and polyols, *N*-methylated compounds and amino acids)					
Metabolite name	**^{1}H NMR δ (multiplicity group)**	**Time point**			
		All	4	10	30
Alanine	1.48 (d)	VD		VD	VD
Acetone	2.22 (s)	VD			
Gamma-amino-*N*-butyrate	2.3 (t)			VD	
Dimethylamine	2.7 (s)			VD	
Trimethylamine-*N*-oxide	3.27 (s)			VD	VD
Sugar*	3.25 (m), 3.43 (m), 3.46 (d)	VD	VD	VD	VD
Tyrosine	7.2 (d)			VD	VD
Tyramine	7.23 (d)			CS	CS
Adenine	8.2 (s)			VD	

Faecal metabolic profiles from infants were then compared at each time point by using OPLS-DA models and it was discovered that the faecal samples differed significantly between vaginally delivered and caesarean section infants up until day 30 and after which there was no significant difference.

At the earliest sample taken (4 days) infant faecal samples separated due to sugars (δ3.25, δ3.43, δ3.46) which were correlated more strongly with infants delivered vaginally. Samples from caesarean section delivered infants had a stronger association with the metabolite tyramine (δ7.23). Vaginally delivered infants' faecal profile, were highly associated with alanine (δ1.48), TMAO (δ3.27), adenine (δ8.2) and dimethylamine (δ2.7). Tyrosine (δ7.2) and gamma-amino-*N*-butyrate (δ2.3) were also related to being delivered vaginally.

4.4.1.5. Nutrition from birth: the effect of breast feeding exclusively on the ^{1}H NMR faecal metabolic profile

An OPLS-regression model of the exclusivity of breast feeding in months was produced on faecal samples taken at 120 days postpartum. The longest period that an infant was exclusively fed breast milk in this study was 12 months but with many infants started to be weaned off milk around 4-6

months this added another confounding factor. Therefore, modelling the data at 4 months gives a better insight to the effect of the exclusivity of breastfeeding in the first few months. This statistical modelling was performed on 138 faecal samples. Individuals which were born at less than 37 weeks gestation age (i.e. preterm infants) and samples with missing nutritional information were not included in the model; discriminatory metabolites associated are summarised in table 6. This model was build using 1 orthogonal component with the final R^2Y and Q^2Y values being 54% and 27% respectively; a relatively robust model.

Faecal metabolic profiles from infants which were fed exclusively breast milk for less time were highly associated with amino acids including butyrate (δ0.91, δ1.57, δ2.16), lysine (δ1.47, δ1.75, δ3.03), propionate (δ1.07, δ2.18), and acetate (δ1.92). These samples were also associated with an unknown metabolite at δ7.16.

Samples from infants which were exclusively fed breast milk for more time were highly associated to 3-hydroxybutyrate (δ1.19), choline (δ3.53, δ4.06) and *N*-acetylglycoprotein fragments (δ2.06).

Table 4-6– List of metabolites assigned in the loadings plot of OPLS regression model of infant's exclusivity to breast fed in months from samples taken at 120 days only.

Metabolite name	**^{1}H NMR δ (multiplicity group)**	**Group observed**
Butyrate	0.91 (m), 1.57 (m), 2.16 (m)	Less exclusivity
Lysine	1.47 (m) 1.75 (m), 3.03 (t)	Less exclusivity
Propionate	1.07 (t), 2.18(m)	Less exclusivity
Acetate	1.92 (s)	Less exclusivity
Unknown 1	7.16 (s)	Less exclusivity
3-hydroxybutyrate	1.19 (d)	More exclusivity
N-acetyl glycoprotein	2.06 (s)	More exclusivity
Choline	3.53 (m), 4.06 (m)	More exclusivity

4.4.1.6. Understanding the effects of early life experiences on the development of the faecal metabolic profile: negative results

Relevant clinical data were collected for each infant at each time point and questionnaires on nutritional was obtained from mothers. This included information on gestational age and weight at birth of the infant; including mode of delivery of the infant and nutritional information: these data have been reported in the literature. Further information on allergies, asthma, colic as well as the developmental condition of the child at two years was also collected. Modelling of faecal water data showed no differences in profile between samples from these classes. (Further information on the model statistical outputs can be found in the supplementary data- Figure S8).

This project is interested in the effect of early life exposures to the correlation to infant BMI status; BMI data were therefore recorded at 1 and 2 years of age. In recent years, the gut microbiota has been indicated in the development of obesity. This development of the gut microbiota community is thought to be influenced by early life exposures including milk composition and the use of antibiotics in early life. The functionality of the complex microbial communities can be further investigated by studying the faecal metabolic profile. Interestingly, modelling the faecal water data showed no significant difference between samples from low and high BMI infants at 1 and 2 years of age. This would suggest that the functionality of the gut microbiome is similar regardless of BMI status of the infants.

Additional information from the mothers was available, including mothers BMI, age, ethnicity, smoking habits as well as antibiotic use during pregnancy and after birth. From the previous work in this thesis, it was notable that urine metabolic profiles at birth from infants can be related to mother's metabolic profile. Therefore, infant faecal metabolic profiles were modelled using maternal classification data to investigate the effect of early life exposures. After modelling these data, no differentiations in metabolic profiles dependant on these classifications were found; this

suggests that faecal samples may not largely be affected by these factors and there are stronger factors that influence the composition of the faecal metabolic profiles.

4.4.2. A TIME SERIES ANALYSIS OF ^{1}H NMR SPECTROSCOPY FAECAL METABOLIC SPECTRA FROM DEVELOPING INFANTS – 4 DAYS POSTPARTUM UNTIL 2 YEARS OF AGE.

After performing multivariate statistics analysis on this data, a time series analysis on the data was also performed on the same data set. As this data set has six time points, it was appropriate to execute this type of analysis to get a better idea of the development of the infant faecal metabolic profile.

4.4.2.1. Investigation of the effect of infants born term vs. preterm on the faecal metabolic profile

Time series analysis showed that there were statistically robust differences in faecal metabolite profiles from infants in the first four time points. Spline graphs from the discriminatory features illustrate at which time point these features differentiate between the two groups and displayed broad trends in behaviour of metabolites over time; infants born term versus preterm (Figure 4-4). Discriminatory metabolites associated with the two classifying groups samples are summarised in table 4-7. These differences include a higher correlation to butyrate (δ0.865, δ0.875, δ0.855; δ1.59, δ1.602, δ1.612) and L1 (δ1.309) in preterm infants at day 4, 10 and 30 time points. At days 10, 30 and 120 days postpartum, preterm infants has elevated relative concentrations of tyramine (δ7.239).

Term infants were correlated with higher concentrations of glucose (δ5.275, δ5.281), tyrosine and choline at days 10, 30 as well as 120 days postpartum. Term infants were highly associated with faecal bile acids (δ0.673, δ0.683) and betaine (δ3.287, δ3.277) at days 10 and 30 postpartum. Interestingly, at day 10, preterm infants correlated highly to an unknown metabolite (δ2.368) but by the time the infants were 120 days old this metabolite was more affiliated with term infants.

Table 4-7: List of discriminatory variables assigned from time series analysis from faecal infants born term (T) versus preterm (P) (*overlapped resonances from glucose, other sugars and polyols, *N*-methylated compounds and amino acids). P-value on the significance of time trajectory of term vs preterm infant samples for variable *x*. Empirical p-value obtained by estimation of the null distribution (no difference) using 1000 permuted random groups, with Benjamini Hochberg FDR correction applied (see page 62).

Variable - δ (observed multiplicity)	*p-value*	Potential ID	Time point			
			4	10	30	120
0.673, 0.683 (s)	0.000999	Bile acid		T	T	T
0.865, 0.875, 0.855 (t)	0.000999	Butyrate	P	P	P	
1.309 (s)	0.000999	Lipid 1 (L1)	P	P	P	
1.592, 1.602, 1.612 (m)	0.000999	Butyrate		P	P	P
1.985 (s)	0.000999	Unknown 1		P	P	
2.369 (m)	0.000999	Unknown 2		P		T
3.287, 3.277 (s)	0.000999	Betaine		T	T	T
3.469 (m)	0.000999	Sugar*		T	T	
5.271, 5.281 (d)	0.000999	Glucose		T	T	T
7.208 (d)	0.000999	Tyrosine			T	T
7.239 (d)	0.000999	Tyramine		P	P	

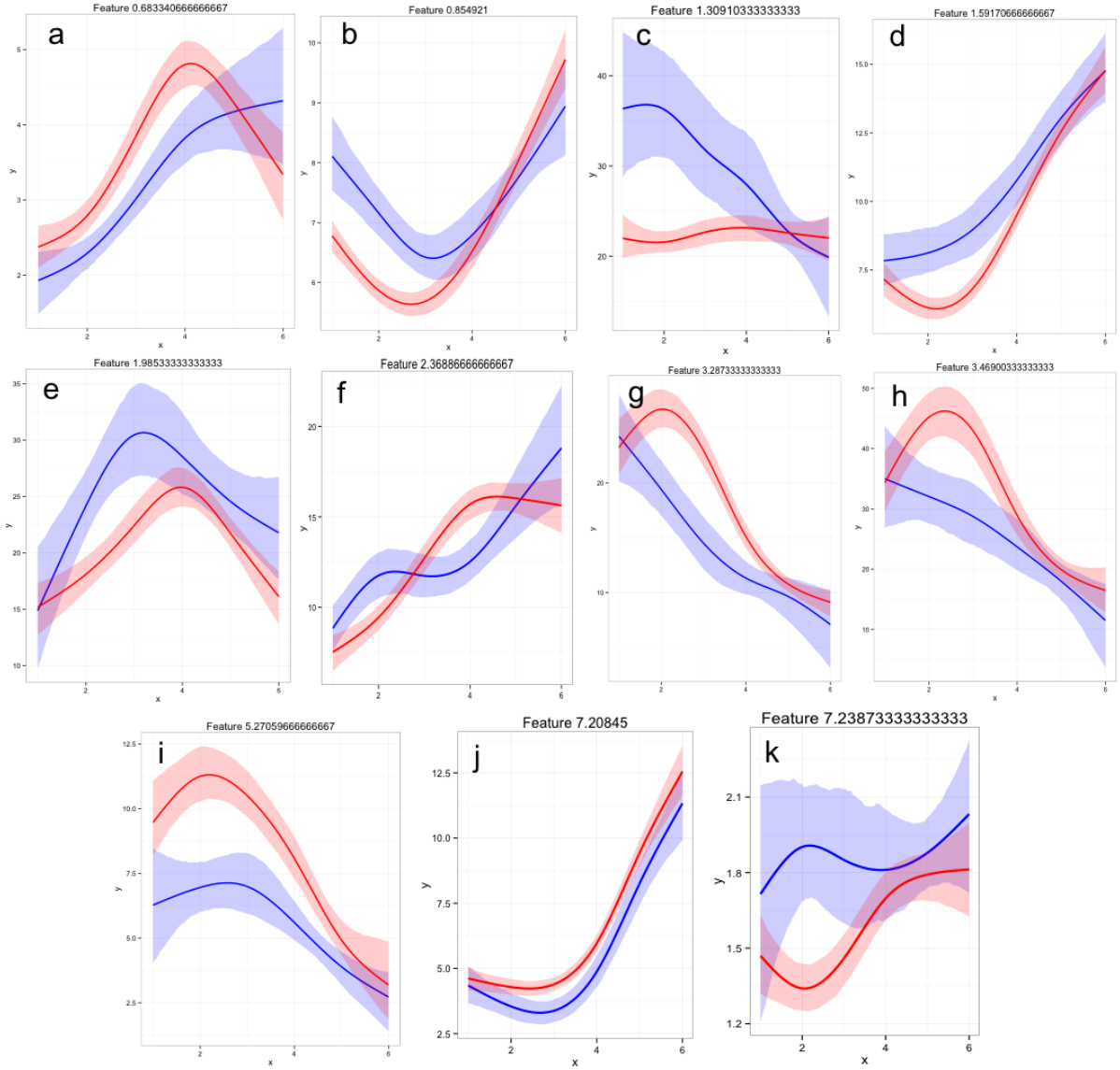


Figure 4-4: Spline graphs of variables taken from ^{1}H NMR spectral binned faecal data from infants born term versus preterm. Line represents the group mean of the feature through time and shaded area represents the 95% confidence interval. Red=Term born infant samples, Blue = Preterm born infant samples. X-axis = time point of which the sample was taken – 1 = Day 4, 2= Day 10, 3 = Day 30, 4 = Day 120, 5 = Day 365, and 6 = Day 730. Y-axis = feature intensity. a) Variable 0.68ppm: Bile Acid, b) variable 0.85ppm: Butyrate, c) variable 1.309ppm: L1, d) variable 1.592ppm: Butyrate, e) variable 1.98ppm: Unknown1, f) variable 2.36ppm: Unknown2, g) variable 3.287ppm:Betaine, h) variable 3.469ppm:Sugar, i) variable 5.27ppm:Glucose, j) variable 7.208ppm: Tyrosine, k) variable 7.238ppm: Tyramine.

4.4.2.2. The effect of the mode of delivery on the faecal metabolic profile in developing infants.

The characterization of the differences in metabolic profiles between infants born vaginally compared to those born by caesarean section using time series analysis was performed. This analysis shows that the main differences between the two groups can be seen at 4, 10 and 30 days' postpartum samples and discriminatory metabolites associated with the two classifying groups samples are summarised in table 4-8. All samples were used in this investigation from all infants, regardless of their gestational age.

Spline graphs from the discriminatory features indicate at which time point these features differentiate between the two groups; infants born caesarean section vs vaginally delivered (Figure 4-5). These differences included a higher correlation to sugars/carbohydrates (δ3.246, δ3.408, δ3.418, δ3.428, δ3.438, δ3.458) and dimethylamine (δ2.712) at these time points in vaginally delivered infants.

Table 4-8: List of discriminatory variables assigned from time series analysis from faecal infants delivered via different modes; vaginally delivered (VD) and caesarean section (CS) (*overlapped resonances from glucose, other sugars and poly-ols, *N*-methylated compounds and amino acids). P-value on the significance of time trajectory of samples from infants with different modes of delivery for variable *x*. Empirical p-value obtained by estimation of the null distribution (no difference) using 1000 permuted random groups, with Benjamini Hochberg FDR correction applied (see page 62).

Variable	**p-value**	**Metabolite ID**	**Higher association at time point**			
			4	10	30	120
1.48 (d)	0.0009	Alanine		VD	VD	VD
2.005 (s)	0.0009	Unknown 1			CS	
2.712 (s)	0.0009	Dimethylamine	VD	VD		
3.257 (s)	0.0009	Trimethylamine-*N*-oxide	VD	VD	VD	
3.246 (m), 3.408 (m), 3.418 (m), 3.428 (m), 3.438 (m), 3.458 (m)	0.0009	Sugar*	VD	VD	VD	
7.198 (d)	0.0009	Tyrosine		VD	VD	VD

At day 10, 30 and 120 days postpartum the metabolites alanine (δ1.48) and tyrosine (δ7.198) were also associated with vaginally delivered infants. At day 30 postpartum, infants born via caesarean section were highly associated with the metabolite found at δ2.005, which I have been unable to be identify to date but which has a chemical shift and multiplicity similar to acetamide.

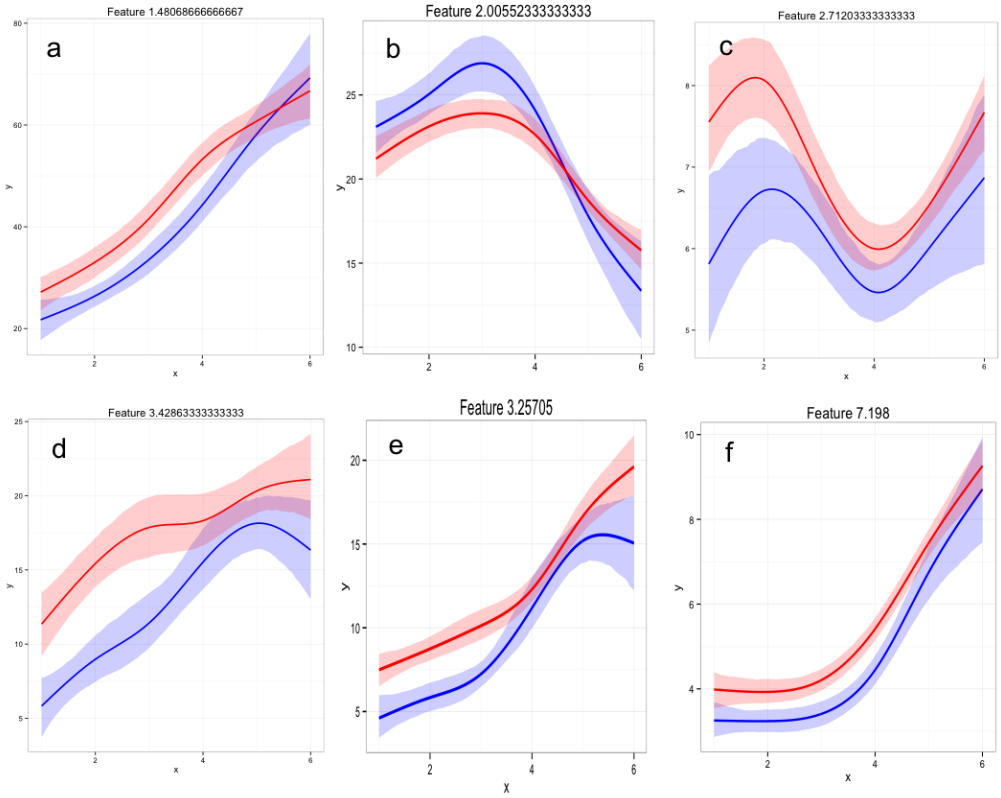


Figure 4-5: Spline graphs of variables taken from ^{1}H NMR spectral binned faecal data from infants born vaginally delivered versus caesarean section. Line represents the group mean of the feature through time and shaded area represents the 95% confidence interval Red=vaginally delivered samples, Blue = caesarean section samples. X-axis = time point of which the sample was taken – 1 = Day 4, 2= Day 10, 3 = Day 30, 4 = Day 120, 5 = Day 365, and 6 = Day 730. a) Variable 1.48ppm: Alanine, b) variable 2.005ppm: Unknown1, c) variable 2.712ppm: Dimethylamine, d) variable 3.24ppm: Sugar, e) variable 3.25ppm: TMA, f) variable 7.198ppm: Tyrosine.

4.4.2.3. The consequence of infant BMI status at 1 and 2 years on the metabolic profile

Unlike the multivariate data analysis, time series analysis demonstrates that there is a difference between infants of different BMI status at 1 year of age. Spline graphs from the discriminatory features displays the time points where these features differentiate between the two groups; high vs low BMI status of the infant at 1 year (Figure 4-7); information gained was categorical of high and low BMI and not the actual BMI number. Discriminatory metabolites associated with the two classifying groups samples are summarised in table 4-9. Infants with a high BMI status at 1 years old, have a relatively higher concentration of trimethylamine (δ2.884) and an unknown metabolite (δ7.925) at 1 years of age and 4-hydroxyphenylacetate (δ7.158) at 2 years old.

Table 4-9: List of discriminatory variables assigned from time series analysis from faecal samples comparing infants BMI status at 1 years of age. P-value on the significance of time trajectory of infant with a BMI high vs low for variable *x*. Empirical p-value obtained by estimation of the null distribution (no difference) using 1000 permuted random groups, with Benjamini Hochberg FDR correction applied (see page 62).

Variable	p-value	Metabolite ID	Time point	
			365	730
2.884	0.00999	Trimethylamine	H	
7.158	0.01198	4-Hydroxyphenylacetate (4HPA)		H
7.925	0.04595	Unknown 1	H	

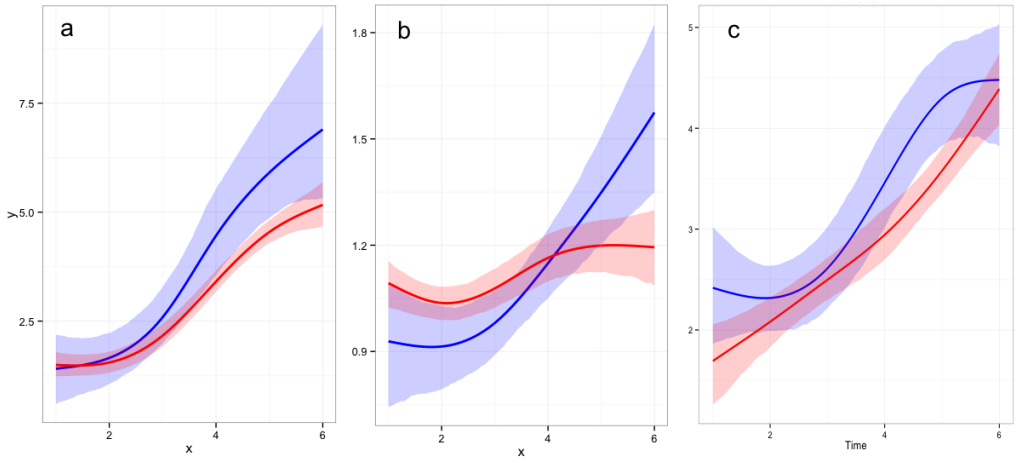


Figure 4-6: Spline graphs of variables taken from ^{1}H NMR spectral binned faecal data comparing BMI status at 1 year – low vs high BMI. Line represents the group mean of the feature through time and shaded area represents the 95% confidence interval. Red=Low BMI infant samples, Blue = High BMI infant samples. X-axis = time point of which the sample was taken – 1 = Day 4, 2= Day 10, 3 = Day 30, 4 = Day 120, 5 = Day 365, and 6 = Day 730. a) Variable 2.88ppm: TMA, b) variable 7.158ppm: 4HPA, c) variable 7.925ppm: Unknown1.

Time series analysis showed that there was a difference between infants of different BMI status' at 2 years of age. Spline graphs from the discriminatory features display at which time point these features differentiate between the two groups; high vs low BMI status of the infant at 2 years postpartum (Figure 4-8). Discriminatory metabolites associated with the two classifying groups samples are summarised in table 4-10.

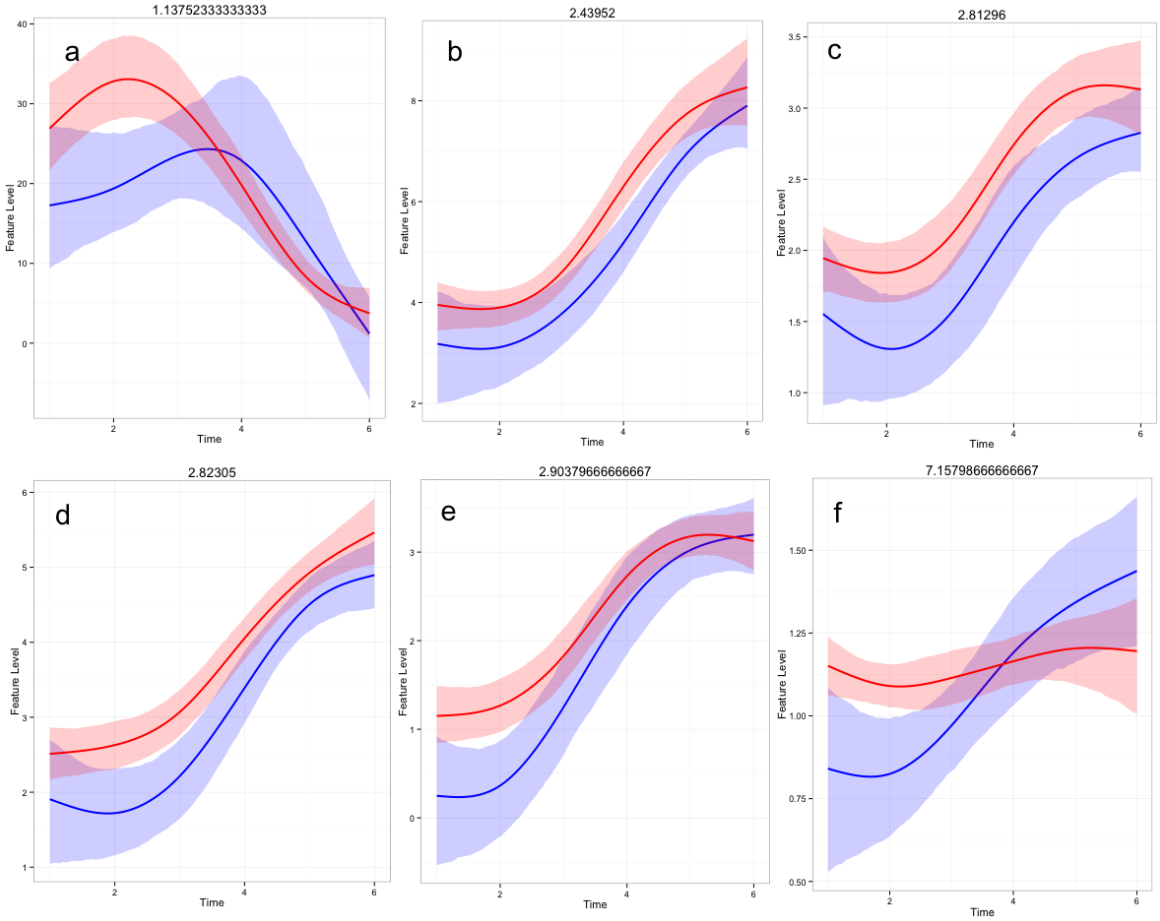


Figure 4-7: Spline graphs of variables taken from ^{1}H NMR spectral binned faecal data comparing BMI status at 2 years of age – low vs high BMI. Line represents the group mean of the feature through time and shaded area represents the 95% confidence interval. Red=Low BMI infant samples, Blue = High BMI infant samples. X-axis = time point of which the sample was taken – 1 = Day 4, 2= Day 10, 3 = Day 30, 4 = Day 120, 5 = Day 365, and 6 = Day 730. a) Variable 1.138ppm: propylene glycol, b) variable 2.439ppm: Glutamine, c) variable 2.81ppm: Aspartate, d) variable 2.823ppm: Aspartate, e) variable 2.904ppm: Unknown1, f) variable 7.158ppm: 4-HPA

Infants with a high BMI status at 2 years old, have a relatively higher concentration of propylene glycol (δ1.138) and 4-hydroxyphenylacetate (δ7.158) compared to infants with a low BMI status at the age of 2 years. Also, an unknown metabolite at δ2.904 was also seen at 10 days postpartum that is associated to a high BMI status at 2 years. When STOCSY was performed on this metabolite it is highly correlated to glutamine at δ2.47.

The metabolite aspartate (δ2.803, δ2.813, δ2.823) was correlated to a low BMI status at three different time points; days 10, 30 and 120 postpartum. Furthermore, samples taken at 120 days have relatively higher concentrations of glutamine (δ2.44) in infants with a lower BMI status at 2 years of age.

Table 4-10: List of discriminatory variables assigned from time series analysis from faecal samples comparing infants BMI status at 2 years of age. P-value on the significance of time trajectory of infants with high vs low BMI for variable *x*. Empirical p-value obtained by estimation of the null distribution (no difference) using 1000 permuted random groups, with Benjamini Hochberg FDR correction applied (see page 62).

Variable ID (δ)	p-value	Metabolite ID	Time point		
			10	30	120
1.138 (d)	0.04795	Propylene glycol	L		
2.44 (m)	0.02797	Glutamine			L
2.803 (dd)	0.04195	Aspartate	L		
2.813 (dd)	0.01098	Aspartate			L
2.823 (dd)	0.00399	Aspartate		L	
2.904 (t)	0.03796	Unknown 1	L		
7.158 (d)	0.03296	4-Hydroxyphenylacetate (4HPA)	L		

4.5. DISCUSSION

Using faecal samples is a relatively novel approach to metabonomic studies; especially for infant studies which are rare. This project has not only been able to provide a comprehensive characterisation of the development of the faecal metabolic profile but also further understand the effect that environmental exposures has. Furthermore, the large scale of this cohort has not previously been performed in infant studies so this gives us an insight into the baseline metabolic profile from birth.

This study has shown that the metabolic profile of infant faecal samples changes rapidly in the first few months of life. In this highly active developmental period, environmental interactions can affect the metabolic profile and potentially health in later life, especially in terms of gestational age at birth, mode of delivery and early nutrition (Malina et al., 2004; Dominguez-Bello et al., 2010; Howie et al 1990; Quigley et al., 2016). Characterisation of the faecal metabolic profile in this sample set has also given further insight into the gut microbiome development and functionality in the first few months of life (Chow et al., 2014). Choline metabolism has been recognised as an important part of development in infants, especially in terms of nutritional changes at 4 months postpartum (Trump et al., 2006).

4.5.1. CHARACTERISATION OF METABOLIC PROFILE FROM FAECES IN DEVELOPING INFANTS OVER TIME FROM BIRTH UNTIL 2 YEARS OF AGE

Infants develop rapidly from birth with consequential adaptation of physiology, which is likely to cause their metabolic profile to dramatically change in the first few months of life. In Chapter 3, it was discovered that urinary metabolic profiles changes dramatically in the first few weeks from birth and can be related to the mother's metabolic profile at the time of birth. In this chapter, faecal metabolic profiles also follow an age-related trajectory. Therefore, this reiterates the importance of standardising age in infant studies, especially when interpreting data for other effects, such as

diseases and co-morbidities. It was observed that carbohydrates, sugars and *N*-acetyl-glycoprotein fragments highly correlated with samples taken at the earlier time points. Breast milk composition consists primarily of sugars (e.g. lactose), carbohydrates (e.g. oligosaccharides), proteins (e.g. casein), lipids (e.g. triglycerides) as well as minerals (e.g. magnesium) (Ballard and Marrow 2013; Andreas et al., 2015).

In the first four to six months of life, the infants in this study were primarily fed milk (breast or formula milk), after which time introduction of solid foods was generally implemented; diet highly contributes to the faecal metabolic profile. Through this period the gut is rabidly establishing a community of bacteria which plays an important role in nutrition by digesting sugars into other products. As the gut microbiome is still being established and stabilising, the sugars from milk aren't being fully broken down, therefore they are excreted in the faeces.

Whereas, faecal water acquired at after 4 months' postpartum correlated with organic acids including, amino acids (alanine, tyrosine, aspartate, phenylalanine), SCFAs (butyrate, propionate and acetate), and BCAAs (valine). Amino acids are important in energy metabolism as they are regularly used in the TCA cycle and intermediates for the synthesis of glucose (Moltu et al 2014). As mentioned previously, the gut microbiome is becoming well-established and many of the bacteria produce short chain fatty acids, essential amino acids and vitamins from the infant's diet and therefore are contributing to the metabolome further (Miller et al., 1996, Louis et al., 2014, Tremaroli et al., 2012).

The greatest change in the faecal metabolic profile can be seen between samples taken at 30 days compared to 120 days postpartum. Infants start to be weaned off breast/formula milk onto solid foods from 4-6 months so this could be the main reason for the dramatic change in faecal profile between these two time points. Furthermore, the gut microbiome can change very quickly during this period and start to become similar to an adult composition (Koenig et al., 2011).

As seen with the previous model, it was observed that sugars and *N*-acetyl-glycoprotein fragments highly correlated with samples taken at 30 days postpartum. However, faecal metabolic profiles at 120 days were correlated to bile acids, 4-cresyl glucuronide, methylamine and an unknown metabolite. This unknown metabolite has two signals at δ1.65 (m) and δ2.24 (t) which is a very similar chemical shift patterns to alpha-aminoadipate; an intermediate of the breakdown of lysine. Further investigation using spike-in experiments would be needed to confirm this ID.

Bile acids have a role in digestion as they emulsify and solubilises fats for intestinal absorption (Houten et al., 2006) as well as repress the growth of microbiome. Furthermore, previous investigations have found that bile acids excreted in the faeces are correlated to the weight of the child; these results are in agreement to these findings as infants at 120 days will have gained a significant amount of weight from 30 days postpartum (Murphy and Signer, 1974).

The metabolite 4-cresyl-glucuronide is a product of protein degradation by the gut microbiome and usually absorbed by the colon. Furthermore, this metabolite is strongly correlated with the transit time of faeces through the gastrointestinal system from urine samples (Roager et al., 2016). From this faecal data, it can be concluded that infants at 120 days postpartum have a higher concentration of protein in their diet but don't fully absorb the by-products from protein breakdown by the gut microbiome (Musso et al., 2010, Dumas et al., 2006). Furthermore, there was no distinguishable difference between 1 and 2 years of age samples, suggesting that at this stage nutrition and the gut microbiome's contribution to metabolism is stable.

4.5.2. CHANGES IN INFANT FAECAL METABOLIC PROFILES RELATED TO EARLY LIFE EXPOSURES

4.5.2.1. Full-term versus preterm (early and late) at birth

Infants born after 37 weeks of gestation are regarded as born at term, whereas infants born earlier are regarded as born preterm; this characterisation can be split again into early (<27 weeks gestation age at birth) and late preterm (32-36 weeks gestation age at birth). From this dataset, I anticipated to gain an insight into the effect of gestational age at birth longitudinally. The difference in early life was further investigated in the subsequent chapter of this thesis (Chapter 5). Furthermore, the model was also modelled at each time point because the faecal metabolic profiles rapidly change over time so it was essential to take away the confounding factor of the data set in terms of time points. From the ^{1}H NMR data through multivariate statistics as well as time series analysis, similar metabolites which are discriminatory between the two groups have been established. From the time series data, the specific trajectory of the metabolites can also be examined thoroughly using Spline graphs.

Butyrate is a principal SCFA that is produced by bacterial fermentation from the breakdown from lactose and recognised here as discriminatory in preterm infant samples. From the time series analysis spline graphs, it is evident that this metabolite has similar time trajectories in term and preterm but at relatively different concentrations. SCFAs are well known to act as a protective agent against infant diarrhoea. However, this metabolite has been implicated in NEC cases in animal models (Pourcyrous et al., 2014, Waligora-Dupriet et al., 2009, Wang et al., 2009). NEC is a serious disease related to preterm infants and the colonisation of pathogenic bacteria in the gut. Unfortunately, there was no data from this project as to whether any of these infants suffered from this disorder during this study.

A lipid (named here as L1 (δ1.3)) in NMR spectra is also highly associated with preterm infants samples (Musacchio et al., 2009). Preterm infants are typically fed high fat parenteral nutrition at the start of life and therefore higher lipid contents in their faeces are expected through malabsorption; preterm infants are more prone to malabsorption due to the immaturity of their gastrointestinal system and the low concentration of bile acids found in preterm infants (Fanaro, 2013). Further evidence for the higher association of L1 in preterm infants can be seen in the spline graphs from the time series analysis as L1 is very high in preterm infants and gradually decreases over time as the infant develops and are subject to less parenteral nutrition.

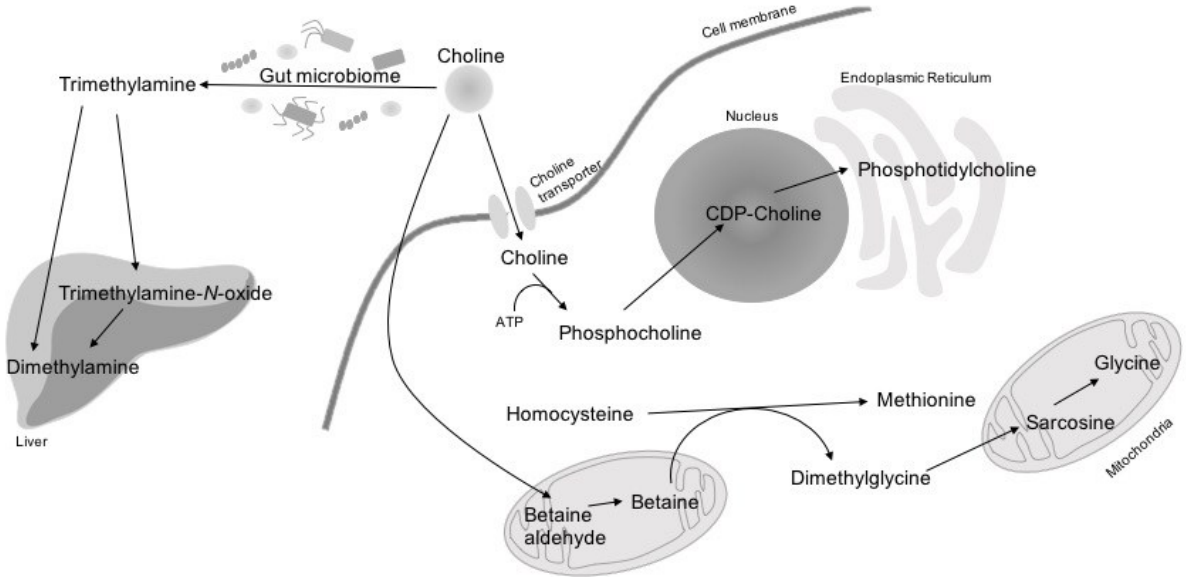


Figure 4-8: Schematic to represent dietary choline biochemical pathway in the human body (Adapted from Cheng et al., 1996)

Term infants are highly associated with choline as well as betaine (choline breakdown product metabolite) from the time series analysis. The human body cannot synthesis choline and relies on exogenous sources from the diet; typically, choline is in high abundance in meat, animal products as well as human breast milk. Choline is broken down by choline oxidase to make betaine aldehyde and broken down further to make betaine by betaine aldehyde dehydrogenase. Betaine is then further metabolised into dimethylglycine (DMG) in another enzymatic reaction when choline is being

converted to glycine. TMA is produced from the conversion of choline by the gut microbiome enzymatic activity. TMA is then absorbed by the body and transported to the liver where TMAO is formed exclusively from the N-oxidation of TMA in the presence of liver FMO enzymes; TMA can also be broken down into DMA by the liver.

Human breast milk provides large amounts of choline, especially in mature milk (220-300 μmol/L) compared to colostrum (~130 μmol/L), whereas formula feed and TPN choline concentrations can vary (~50-700 μmol/L) (Ilcol et al., 2005; Caudill 2010; Cheng et al., 1996). Choline is a precursor for the biosynthesis of phosphatidylcholine which is essential to make cell membranes and therefore, neonates need large amounts of choline to sustain growth (Holmes-McNary et al., 1996). Term infants are also highly associated with *N,N,N*-trimethyl-lysine; this metabolite is a methylated derivative of the amino acid lysine. It is an integral part of carnitine biosynthesis and essential for energy metabolism (Vaz et al., 2002; Servillo et al., 2014).

Separating the data set and modelling the data by time point independently, gives a better insight into the changing development of the faecal metabolic profile. Faecal water NMR profiles of preterm infants are highly associated with 3'-fucosyllactose, a common breast milk oligosaccharide and a substrate that the infant doesn't digest in the stomach (Practico et al., 2013). This metabolite reaches the large intestine where the gut microbiome utilises it as a source of nutrition. Therefore, this could suggest the term infants' gut microbiome have a larger diversity than the preterm infants and fully utilising this source of nutrition and therefore not being excreted by the term infants too (Cong et al., 2016).

Using time series analysis, the spline graph for tyramine is very interesting. With similar relative concentrations of tyramine at birth for both term and preterm born infants, this metabolite decreases in term samples in the first few days of life and then rapidly increases after 10 days postpartum. Whereas in preterm infant samples tyramine stays relatively stable throughout the time points. This metabolite is a derivative of the common amino acid, tyrosine; the conversion of this

metabolite is linked to lactic acid bacteria, a common bacterial species found in the neonatal gut (Marcobal et al., 2013). It would be interesting to correlate this information to bacterial profiles and know the abundance of lactic bacteria (Johansson et al., 2005).

Term infants have higher concentrations of faecal bile acids than preterm infants from 4 to 120 days postpartum, which has been seen previously (Moltu et al., 2014). This result reiterates the baseline characteristics seen in term infant faecal profiles compared to preterm. From the spline graph for this variable, the increase of bile acid at 120 days postpartum could be related to the shift in diet that infants start to experience when they are weaned off milk and fed solid foods.

In term infant samples at 120 days postpartum, the metabolic profile has changed considerably compared to 30 days postpartum, which is most likely to do with the weaning of the infant. Term infants at 120 days postpartum are still associated with choline however. Also at this time point, term infants had a higher correlation to amino acids, including alanine, tyrosine, phenylalanine and leucine, which is likely to be related to their change in diet and the increase production of energy for growth.

Term infants were related also to succinate at this time point which is involved in the citric acid cycle to produces energy for cells. This metabolite has previously been found to be related to age in urine, with stronger correlations in infants less than one years of age compared to toddlers (Gu et al., 2009), therefore this metabolite may be higher in term infants due to their relative size at this time point compared to preterm infants. Furthermore, succinate can be broken down from glucose by the gut microbiome, further suggesting that the gut microbiome community is more diverse in term infants (Jurtskuk et al., 1996).

Faecal samples taken at days 365 and 730 postpartum, do not show any separation in samples from infants born term or preterm. This suggests that being born term or preterm on the metabolic profile doesn't carry on once the infant is older.

4.5.2.2. Birth gestational age (GA) and birth weight

As previously discussed, birth weight and birth gestational age (GA) could be related, and from this data this correlation is reflected in the similarity of the metabolites associated with each class. Modelling the data using a regression analysis was performed to gain a greater insight into the potential subtle differences in the faecal metabolic profile, which may be lost from a discriminatory analysis (term *vs* preterm). Faecal metabolic profiles were analysed as a whole data set which included all time points, as well as analysing the time points individually, using a regression model (OPLS).

When modelling all the time points, infants born at a lower birth GA and weight had relatively higher concentrations of ethylene glycol. As described in chapter 3, this metabolite has been recognised as a contaminant from nappies, however it can also come from other sources including drug excipients. Infants who are born at a lower GA or weight are more likely to have complications, related to being underdeveloped from birth, and consequently are more likely to be treated with a variety of drugs. Furthermore, infants born at a lower birth GA and weight are related L1 (δ1.31); a tentatively assigned fatty acid chain of a triacylglyceride. Infants born at a lower GA are more likely to be being fed a diet rich in fatty acids (including linoleic acid) as well as other nutrients (amino acids, glucose, vitamins and minerals) (Gutcher et al., 1991, Meyers et al., 2013).

Whereas, only infants with a lower birth GA and not birth weight were positively correlated to butyrate (δ0.86); this metabolite was also correlated to infants born preterm. Butyrate is a common source of energy for colonic epithelial cells and has a role in the proliferation of these cells as well as the maintenance of the intestinal barrier function (Neu et al., 2007, Hague et al., 1996). Formula fed infants produce more butyrate compared to breast fed infants and as these infants are more likely to be preterm this result is not unusual (Sanderson and Naik, 2000; Lee et al., 1972). As mentioned before, this metabolite has been implicated in NEC (Pourcyrous et al., 2014; Waligora-Dupriet et al., 2009) and these results suggest that infants with a higher birth weight irrespective of their

gestational age may not be as likely to contract NEC (Gephart et al., 2013). This information could be useful for the management of preterm infants and further understanding the aetiology of NEC.

Conversely, choline and *N,N,N*-trimethyl-lysine was at a relatively higher concentration in samples from infants born with a higher GA and birth weight as well as glucose and other sugars. These discriminatory metabolites are the same as those seen in term infants, when modelling infants as the two discriminatory classes; term or preterm.

By modelling the data from each time point separately, samples taken at 4, 10 and 30 days with a lower gestational age and a lower birth weight were positively correlated to 3'-fucosyllactose, a common breast milk oligosaccharide. It is believed that oligosaccharides are resistant to enzyme hydrolysis in the stomach so that they are able to reach the large intestine where they serve as substrates for bacterial metabolism (Marcobal et al., 2012). This would suggest that the gut microbiome from this group are not utilising this compound fully so it is being excreted. Recently, Andreas et al., (2016) investigated the role of specific oligosaccharides species on Streptococcus B (GBS) infection in infants. This is important as GBS infection is the leading cause of infection in the first month of age. This study concluded that the 3'-fucosyllatose is inversely correlated to the abundance of GBS in breast milk as well as infant gut microbiome (Andreas et al., 2016); this is due to HMOs stopping the adherence of pathogenic bacteria to the intestinal epithelium as well as promoting commensal bacteria by acting as a nutritional source.

Infants born at a higher birth weight are related to bile acids at 4 and 10 days postpartum, as previously been described these samples are related to weight rather than the GA of the infant at birth (Murphy and Signer 1974).

When analysing the data as a whole it was apparent that at 120 days postpartum the metabolic profile changes significantly compared to samples taken during the neonatal period (≤30 days postpartum). When regressing birth weight and birth GA at 120 days postpartum, samples from

lower birth GA and weight separate due to higher correlation to sugars and carbohydrates; however, this does not include the milk oligosaccharide, 3'-fucosyllactose. These infants typically won't be fed as much breast milk at this stage and are either on formula feed or weaned onto solid foods. Samples from higher birth GA and weight separate due to amino acids including alanine, valine, tyrosine and leucine as well as succinate. These discriminatory metabolites are comparable in term infants, when modelling infants as the two discriminatory classes (term *vs* preterm) rather than a regression.

Comparable to the previous model, faecal samples taken at days 365 and 730 postpartum, do not show any separation in samples from infants at any birth weight or gestational age. This suggests that the differential effect of birth weight and gestational age on the metabolic profile doesn't carry on once the infant is older.

4.5.2.3. Mode of delivery – Vaginally Delivered (VD) vs. Caesarean Section (CS)

From the ^{1}H NMR data through multivariate statistics as well as time series analysis, similar metabolites which were discriminatory between the two groups have been established. From the time series data, the differential trajectories of the metabolite profiles associated with mode of birth can also be examined.

Infants can be delivered through the vaginal canal (VD) or by caesarean section (CS); these actions are different mechanically as well as biologically. One very important difference between these infants is that they are thought to have different microbiota colonising their guts initially depending on the mode of delivery. Vaginally born infants are colonised by bacteria which are typically found in the vaginal canal (e.g. *lactobacillus*), whereas infants which are born through caesarean section are colonised by bacteria typically found on skin or within the hospital environment (e.g. *Staphylococcus*) (Dominguez-Bello et al., 2010, Matamoros et al., 2013). Furthermore, in the first few days after birth, the gastrointestinal system is becoming increasingly more anaerobic, thus

rapidly affecting the community (Solis et al., 2010). By analysing faecal metabolic profiles, it gives an insight into the functionality of the gut microbiome and whether differences can also be seen in infants born through different modes.

Infants which were delivered vaginally (VD) have relatively elevated concentrations of alanine when all the time points were modelled as well as at 10 and 30 days postpartum. This metabolite is a common amino acid; mainly used to produce energy from protein degradation (Rodwell, 2012). Acetone was also associated with this group; this is a ketone body, which is broken down in the liver to make energy for muscles, kidneys and brain (Bronk, 1999; Raman et al., 2013). This would suggest that energy production is more efficient in VD infants compared to CS.

By modelling the data separately by time point, samples at day 10 and 30 postpartum from infants born through different modes of delivery separated clearly. Infants born through caesarean section correlate highly with tyramine. As previously mentions in the term versus preterm analysis, this metabolite is related to blood pressure. It has been found that delivery through CS causes a higher risk to high blood pressure in later life (Horta et al., 2013); therefore, this metabolite may be a factor towards this risk. Furthermore, tyramine is a derivative of the common amino acid, tyrosine; this conversion is linked to lactic acid bacteria (Pessione, 2012,). Lactic acid bacteria are commonly found in the infant gut microbiome as breast milk is a significant source for these bacteria to colonise the gut (Martin et al., 2003). Infants born vaginally already have an abundance of *Lactobacillus* in their guts whereas CS infants are typically colonised by *Staphylococcus* bacteria (Dominguez-Bello et al., 2010). This result suggests that infants born through CS aren't introduced to lactic acid bacteria until they are breast fed and this influx of bacteria can then utilise tyrosine and convert it into tyramine. Interestingly, infants born vaginally positively correlated to tyrosine at this time; the amino acid precursor to tyramine. Schulpis et al., (2008) investigated the plasma from mothers and umbilical blood who delivered their babies through VD and CS. They concluded that VD increased tyramine

found in plasma from mothers as well as umbilical cord blood and attributed it to the stress of delivery (Schulpis et al., 2008).

Dimethylamine (DMA) and trimethylamine-*N*-oxide (TMAO) are related to the degradation of choline from the diet by the gut microbiome (Zeisel et al., 1989). The bacteria in our gut break down choline into trimethylamine (TMA) and DMA. TMA is absorbed and transported to our livers where it is further metabolised into TMAO (Romano et al., 2015). The association of these metabolites with VD infants may be due to the excess of choline from breast milk of these infants. Interestingly, from the time series analysis, the spline graph for dimethylamine is high at the beginning of life, then drops off at 4 months before increasing again at 1-2 years; this result is similar to previous findings (Foxall et al., 1995). The increase of choline concentration in breast milk over the first few days and then the introduction of solid foods further changes the composition of the gut microbiome and therefore the functionality which can be recognised in the faecal profiles (Holmes et al., 2000).

At 10 days postpartum infants VD had relatively higher amounts of γ-aminobutyrate (GABA), this compound is the product from glutamate break down by the *Lactobacilli* and *Bifidiobacteria* (Barrett et al., 2012; Li et al., 2010). These species of bacteria have been shown to be present at higher levels in infants born via VD so this result is not surprising (Dominguez-Bello et al., 2010). Although, GABA is a well-known neurotransmitter and elevated levels have been implicated in neurological disorders including autism spectrum disorders (Dhossche et al., 2002).

4.5.3. Differences in metabolic profiles of stool associated to nutrition in infants exclusively breast fed for different periods of time

Breastfeeding infants is recommended due to the complex mixture human oligosaccharides present in the milk as well as the immunological benefit for the child; breast milk is rich with immunoglobulin to protect the infant as well as help to mature their immune system (Andreas et al., 2015). Both of these functions are difficult to replicate in infant formula feed. To investigate the specific metabolic

signature associated with diet-induced development, faecal metabolic profiles were studied at samples taken at 120 days postpartum. Information was collected from mothers how long, in months, were infants exclusively breast fed (maximum exclusive breastfeeding for 12 months) but with many infants started to be weaned off milk around 4-6 months this added another confounding factor. Therefore, modelling the data at 4 months gives a better insight to the effect of the exclusivity of breastfeeding in the first few months.

Breast milk differs from infant formula feed as it contains sialylated and fucosylated oligosaccharides (human milk oligosaccharides), which acts as prebiotics for the gut microbiome (Marcobel 2012; Tannock 2013; Smilowitz 2014). It therefore contains *N*-acetyl glycoproteins and choline, which corresponds to the results seen from infants exclusively breast fed for 4 months. Choline has been found to be critical during fetal development and is important for normal growth and neuro-development in neonates (Zeisel 2006; Ozarda 2014; Ilcol 2005). This reiterates the need to feed infants breast milk from birth, especially if infants being formula fed are not gaining enough choline for normal development, which these results may be reflecting.

Infants that weren't exclusively fed breast milk were positively correlated to SCFAs including butyrate, acetate and propionate. These fatty acids are bacterially generated by the fermentation of polysaccharides and infants who are formula fed have a more diverse bacterial community including *Bacteroides* and *Clostridium* species, compared to breast fed infants, who predominantly contain *Bifidiobacterium* (Knol et al., 2005). In agreement with previous findings, it has been found that butyrate and propionate are found at a higher concentration in formula fed infants compared to breast fed infants (Edwards et al., 1994; Edwards et al., 2002). Additionally, acetate is usually associated to breast fed infants (Knol et al., 2004). However, as this project is investigating the faecal metabolic profile it would indicate that these molecules are not being absorbed by the infant and so not introduced into the body systemically. Therefore, it could be postulated that acetate may be being produced by breast fed infants but they may be fully absorbing it and excreting less readily in

their faeces. If we would be able to profile urine samples from the same infants this would give an indication of the metabolites which have entered the body systemically and potentially have interactions with their own metabolism.

Furthermore, formula feed contains an excess of protein compared to breast milk to make up the amino acid differences and therefore these molecules are in excess for the formula feed infant (Chow et al., 2014). In animal models, it has been suggested that this excessive protein intake could compromise the intestinal barrier function, alter the colonic microbiota as well as affect early immune development (Boudry et al., 2013). Additionally, high protein infant formula feeds also contribute to obesity risk later in life (Lifschitz et al., 2015).

4.5.4. Time series analysis of ^{1}H NMR faecal metabolic profiles from infants classed as high BMI at 1 year and 2 years old.

Obesity is among the major current health problems, which are increasing rapidly worldwide. Obesity is associated with a heightened risk of Western lifestyle diseases such as type 2 diabetes, cardiovascular disease risk, and sleep apnoea, as well as also increases the mortality risk (Jess, 2015; Musso et al., 2011). Excessive weight increments during the first months of life have been associated with a high risk of obesity development at 3 years of age as well as in later life (Rodriguez et al., 2015; Druet et al., 2011; Monterio and Victoria 2004). This project is interested in the effect of early life exposures to the correlation with infant BMI status; BMI data were therefore recorded at 1 and 2 years of age.

From the ^{1}H NMR data, the multivariate model (OPLS-DA) indicated that there was no significant difference in faecal metabolic profiles from infants with a high versus a low BMI. However, from the time series analysis on the same ^{1}H NMR data, there is a significant difference between metabolite profiles between infants with high versus low BMI status. However, the difference between the different BMI status' was very small as can be seen in Figures 4-7 and 4-8.

At 1 year of age, infants with a high BMI status at this age were associated with the metabolite, trimethylamine (TMA). TMA is a product of choline metabolism by the gut microbiome, where it is usually absorbed by the body to convert it into trimethylamine-*N*-oxide (TMAO) in the liver (Figure 9). From the spline graph, it is evident that this metabolite is increasing as the infant ages, showing the increased efficiency of the gut microbiome metabolising choline products. TMA and TMAO in urine and plasma have been connected to a higher risk of cardiovascular disease and related to the development of diabetes and obesity (Wang et al., 2011). Furthermore, Zhang et al., (2015) noted a decrease in TMA in urine from obese children on a non-digestible carbohydrate intervention diet indicating that this metabolite is a factor towards developing obesity in infants (Zhang et al., 2015).

4-Hydroxyphenylacetate is a microbial catabolism product of tyrosine and was associated with infants with a high BMI status at 1 year of age (Holmes et al., 2011). From the spline graph, this metabolite increased rapidly after the first month of life until 2 years compared to infants with a low BMI status where it has a relatively stable trajectory through time. 4-Hydroxyphenylacetate has previously been noted to be elevated in biofluids associated with GI conditions in children, including small bowel disease (Chalmers et al., 1979). Protein degradation products suggest that the gut microbiome from these infants is has become distinct after the first month of life and is utilising tyrosine for energy thus producing 4-hydroxyphenylacetate. Interestingly, this metabolite is present in samples taken at 10 days postpartum, and is at a higher concentration in infant with a low BMI status at 2 years of age. When observing the spline graph, it shows that this metabolite is at a low concentration at birth and increases exponentially over time in infants with a high BMI whereas remained relatively stable for infants with a low BMI status at two years. This implies that high BM status infants at 2 years of age may have a distinctive gut microbiome altering their metabolic profiles. Further investigation into which bacterial species associated to this metabolite would be very interesting.

Infants with a low BMI status at 2 years of age are connected to aspartate at 10, 30 and 120 days postpartum with the spline graphs indicating that the trajectory is similar for each group but increased in these infants. Aspartate is a non-essential amino acid that is generated in the urea cycle by citrulline and ornithine (Rodwell, 2012). From the time series analysis, aspartate was also significantly different with a p value of 0.04 from samples taken at 30 days postpartum indicating it was not highly significantly different between the two groups at this time point.

Furthermore, at 120 days postpartum, samples from infants at a low BMI status at 2 years of age are positively correlated to glutamine. Glutamine an amino acid found at high concentrations in human breast milk and increases in concentration in breast milk with time (Agostoni et al., 2000). This may indicate that infants that were still being fed breast milk at 120 days postpartum were more likely to have a lower BMI once they reach two years of age. This result has previously been observed in infants who are breastfed for longer are found to be leaner than those who are weaned earlier (Burke et al., 2005, Weng et al., 2012).

4.6. CONCLUSION

There is a clear age related variability in faecal metabolic profile which has been characterised in this project. Faecal metabonomics is fast becoming a tool to help us understand interactions between human health, nutrition and the mutualistic relationship of the gut microbiome. This interaction is especially important throughout the infant developmental period as it could have an impact on later life health.

Changes in nutrition dramatically alters the faecal metabolic profile at 4 months postpartum and choline metabolism has been recognised as an important function in infant metabolism. Furthermore, the gut microbiome functionality in energy metabolism and metabolic pathways is very interesting and opens up new ideas for targeted investigation in infant development.

Gestational age at birth is recognised as a contributor the metabolic profile as well as the mode of delivery; especially the potential altered functionality of the gut microbiome in these infants.

This cohort of infants has progressed the understanding of the development of the faecal metabolic profile. This foundation will help to determine the relative strengths in future studies especially in terms of gestational age at birth, early nutrition, mode of delivery, the role of the gut microbiome and the risk of obesity later in life.

CHAPTER 5

METABOLIC AND MICROBIOMIC PROFILING OF INFANTS BORN TERM VERSUS PRETERM

5.1. INTRODUCTION

Preterm infants are generally born immature as different systems develop at different times of the gestational period. Preterm and low birth weight (LBW) babies that survive through the neonatal period face the possibility of a lifetime of different disorders and disabilities (Howson et al., 2012; Moster et al., 2008). Preterm birth may predispose children to the development of the symptom complex referred to as metabolic syndrome in later life. Metabolic syndrome can be defined as a group of risk factors that raise the possibility of heart disease, stroke and other health problems that affect the blood vessels (Alberti et al., 2005; Despres et al., 2006; Reaven, 1988). Metabolic syndrome starts with early insulin resistance, with the consequent development of type II diabetes, abdominal obesity, and hypertension (Abitbol et al., 2012).

Furthermore, early life environmental influences, including gestational age, the mode of delivery, nutrition (e.g. maternal or formula milk), clinical interventions (e.g. antibiotics), environmental exposures can affect development of infant development as well as the composition of the gut

microbiome (Nicholson et al., 2012; Penders et al., 2006). For many preterm infants, they have a perturbed development of the gut microbiome due to these different factors (Cong et al., 2016).

In studies investigating preterm infants compared with full term infants, an altered urinary metabolome has been discovered in preterm infants, with increased relative concentrations of citrate found in preterm and creatinine in term infants (Hyde et al., 2010). Furthermore, metabolites related to renal function during the neonatal period have been recognised as different in preterm infants, including 1-*N*-methylnicotinamide and myo-inositol (Foxall et al., 1995; Trump et al., 2006).

As previously noted in chapter 4, faecal metabolic profiling is still a novel application for metabonomic studies. Using this biofluid will help us to further understand the relationship between human health, nutrition and the developing gut microbiome. Currently there is limited knowledge and research in the infant faecal metabolome (Chow et al., 2014). This study complements the previous chapter by specifically looking at term and preterm infants metabolic profiles but at specific time points; at birth, term equivalent age and 3 months postpartum.

Postnatal age is highly related to the changes in the metabolic profiles as well as the gut microbiome, due to the rapid development of the body after birth. Exposures to different environmental factors add to the importance of the neonatal period for development and health in later life. Therefore, the understanding of different exposures in early life can help to predict or prevent later disorders.

5.2. AIM

The aims of this chapter are to:

- Characterise the dynamic perturbations of urinary and faecal metabolites in infants born preterm or term by ^{1}H NMR spectroscopy and UPLC-MS in conjunction with multivariate data analyses.

- Understand the variation in urine and faecal water metabolic profiles in infants through the first few weeks of life by ^{1}H NMR and UPLC-MS.

- To observe changes in the gut microbiota community during the course of development from birth in infants born term or preterm.

5.3. MATERIALS AND METHODS

5.3.1. SAMPLE SET INFORMATION

Samples of biofluids were taken from infants shortly after birth and during the first few weeks of life. The infants in this study were born at either term (gestational age (GA) 37-41 weeks) or preterm (GA 24-36 weeks). This study includes urine (n=278) and stool (n=308) samples from about 150 infants (Figure 5-1 and Table 5-1). Biofluid samples taken from preterm infants were taken from birth and subsequently at weekly intervals until they were discharged from the neonatal unit. Whereas, urine and stool samples from term infants were collected predominantly at birth or in the first two weeks of life, and over the first 3 months of life, with the majority taken around 3 months postpartum. Information corresponding to birth gestational age, day of sample postpartum and type of feeding was also recorded at the time of sampling.

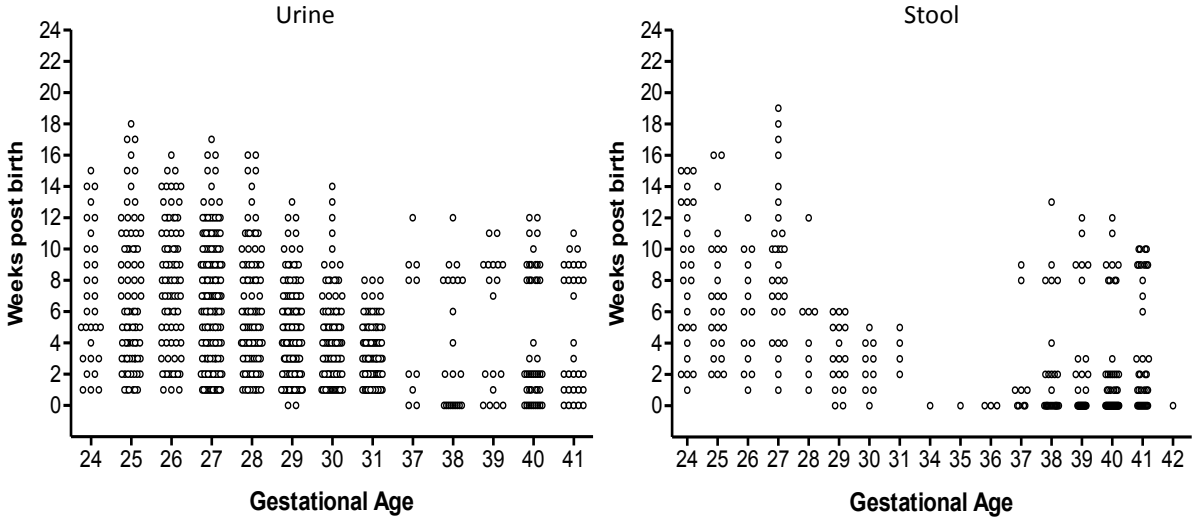


Figure 5-1: Schematic representation of urine and faecal samples analysed in chapter 5 for metabolic phenotyping and metagenomic analysis. Showing gestational age as well as date at which sample was taken in weeks postpartum

Table 5-1: Clinical data information for Term *vs.* Preterm sample set. Urine and Stool samples analysed using ^{1}H NMR and LC-MS. Smaller batch of stool sampled for 16S MiSeq analysis due to sample volume. Data is expressed as the median (IQR) or as a percentage of subjects.

Clinical data	Urine	Stool	16S MiSeq
Number of samples	278	308	296
Number of infants	109	151	147
Time points range (days)	37 (0-123)	21 (0-134)	21 (0-134)
Birth samples (%)	11.10%	7.48%	6%
Gender (% male)	49.10%	47.90%	47.47%
Birth gestational age (range)	30.7 (24.1-42)	38.3 (24.1-42)	38.5 (24.1-42.2)
Preterm samples (%)	54.50%	44.48%	43.77%
Weight median (g)	2770 (590-4720)	3050 (590-5820)	3050 (590-5820)
Caesarean section (%)	19.30%	23.40%	23.90%
Breast fed infants	46.60%	26.30%	26.60%

Urine and stool samples were collected from infants born at term and preterm as part of an on-going study at the Chelsea and Westminster Hospital (Ethics number: REC10H07135). Urine and stool was stored in Chelsea and Westminster hospital in -80°C freezers and transferred to the South Kensington Campus using dry ice to keep them frozen. Samples were then stored in -80°C freezers in the Sir Alexander Fleming building until preparation for metabolic profiling was conducted.

5.3.2. URINE PREPARATION FOR METABOLIC PROFILING (^{1}H NMR AND UPLC-MS)

Urine samples were thawed and centrifuged at 16000 *xg* for 10 minutes at 4°C. A volume of 540µl of urine was mixed with 60µl of phosphate buffer solution at pH 7.4 (Potassium Dihydrogen Phosphate (KH_2PO_4), 3-Trimethylsilyl Propionic Acid-d4 Acid Sodium Salt (TSP), Sodium Azide (NaN_3)), containing TSP (0.5mM) and sodium azide (3mM). Samples were transferred to 5mm NMR tubes and spectra acquired using a 600MHz ^{1}H NMR spectroscopy.

For UPLC-MS metabolic profiling, urine samples were thawed and centrifuged at 16000 *xg* for 10 minutes at 4°C. Samples were transferred to 96 well plates for analysis by UPLC-MS. A pooled sample from all urine extracts were created for quality control (QC) purposes.

5.3.3. FAECAL WATER PREPARATION FOR METABOLIC PROFILING (^{1}H NMR & UPLC-MS)

Faecal samples were weighed out (100mg) and transferred to micro-vials, which contained 0.1mm zirconium beads whilst frozen. They were subject to two freeze thaw cycles before having 250µl of ACN:H_2O added to the micro-vials. These vials were homogenised using a bead beater for 10 seconds and centrifuged at 16,000 *xg* for 10 mins, at 4°C. The supernatant was taken from the vial and placed in pre-washed spin columns and again centrifuged for 20 mins at 16,000 *xg*, at 4°C. The faecal water was separated into two aliquots; one for NMR analysis and one for UPLC-MS analysis. The NMR aliquot was evaporated under nitrogen gas for 30 mins to remove the organic solvent. Once evaporated, these samples were reconstituted into 540µl D_2O and 60µl phosphate buffer, vortexed and sonicated for 5 minutes. The faecal water samples were transferred into 5mm NMR tubes for spectra to be acquired using a 600MHz ^{1}H NMR spectroscopy.

The second aliquot was used for targeted bile acid profiling using UPLC-MS analysis; the faecal water samples were diluted 1:1 sample to isopropanol (IPA) respectively. It was vortexed for 1 minute, frozen for one hour and centrifuged for one hour at 16,000 *xg*, and the supernatant was transferred into glass vials.

The faecal pellet obtained after the first centrifugation step during faecal water extraction, were subject to DNA extractions for gut microbiome investigation; this process is further explained in section 3.6 of this chapter.

5.3.4. NMR SPECTROSCOPY OF BIOFLUIDS

Please refer to Chapter 3, section 3.3.5 for details (Materials and Methods).

5.3.5. ULTRA-PERFORMANCE LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY (UPLC-MS)

5.3.6. Hydrophilic Interaction Chromatography – Mass Spectrometry (HILIC-MS)

Please refer to Chapter 3, section 3.3.6.2 for details (Materials and Methods).

5.3.7. Targeted Bile Acid Profiling using UPLC-MS

An acquity UPLC BEH C8 (2.1 × 100 mm, 1.7 μm) column (Waters Corp, USA) was used with the column temperature set at 60 °C. Mobile phase A consisted of water acetonitrile (ACN)/ (90:10) and mobile phase B ACN/IPA (50:50). In mobile phase A solution ammonium acetate was diluted to 1 mM final concentration and 0.35 mM acetic acid. The elution gradient was set as follows: 90% A (0.0−9.25 min; 0.6 mL/min), 90−65% A (9.25−11.5 min; 0.6 mL/min), 65−15% A (11.5−11.8 min; 0.65 mL/min), 15--0% A (11.8—12.4min; 0.8−1 mL/min), 0 - 45% A (12.45-12.5 min; 0.85 mL/min), 45 - 90% A (12.5-12.6 min; 0.85 mL/min), 90% A (12.6−12.7 min; 0.8 mL/min), 90% A (12.7−12.8 min; 0.7 mL/min), 90% A (12.7-15 min; 0.6 mL/min). An injection volume of 1 μL was used for both positive and negative ionization polarity modes. The auto-sampler was set to operate at 4°C. Mass spectrometry was performed using a Xevo G2 Q-TOF (Waters MS Technologies Ltd., UK) with an electrospray ionization (ESI) source. Finally, the same QC strategy was followed as for this analysis with H_2O experimental blanks and a pooled infant faecal water sample being injected at the beginning and end of the acquisition. Furthermore, bile acid standards mixes were also acquired for reference.

5.3.8. GUT MICROBIOME ANALYSIS FROM FAECAL SAMPLES USING 16S rRNA GENE SEQUENCING

Pellets taken from faecal samples, obtained from faecal water samples extracted previously, were subject to DNA extraction. Isolation and purification of DNA from faecal samples was performed by

PowerFecal® DNA Isolation Kit (MO-BIO Laboratories, California USA). This process involves the breakdown of cells using bead homogenisation as well as chemical disruption to cell walls, to ensure all DNA is extracted from cells. DNA is isolated through binding to a silica spin column and further washing to remove any substances that may interfere with PCR before eventually being eluted from the column. This eluted DNA was used for next generation sequencing.

After DNA extraction of faecal samples, Polymerase Chain Reaction (PCR) was performed. The PCR mixture included 0.25 µl of *Taq* polymerase, 2.5 µl 10X buffer, 0.25 µl BSA, 0.5 µl of the universal set of primers for the variable region V3 (515F, 806R), 0.2 µl of dNTPs, 1 µl of the extracted DNA sample and making up the mixture to 25 µl with nuclease free water. PCR methods were performed by an initial denaturation step at 95°C for 3 minutes, followed by 30 cycles each consisting of a denaturation step at 90°C for 30 seconds, an annealing step at 58 to 65°C for 30 seconds and an extension step at 72°C for 45 seconds. A final extension step of 10 minutes at 72°C was performed to ensure complete amplification of all the DNA fragments. PCR products were analysed by electrophoresis in agarose gels containing ethidium bromide at a concentration of 2% w/v.

To sequence the purified DNA samples, they were sent to Research and Testing Laboratory (USA) where sequencing was performed on Illumina MiSeq. More details of this technique can be found in the Methods section (chapter 2).

5.3.9. MULTIVARIATE DATA ANALYSIS

5.3.9.1. ^{1}H NMR data processing

Please refer to Chapter 3, section 3.3.7.1 for details (Materials and Methods) Urine and faecal water data sets were treated separately for processing.

5.3.9.2. Data Extraction for global and targeted UPLC-MS

Data collected following HILIC-MS analysis (both positive and negative) were subjected to peak-picking, grouping and alignment using MarkerLynx XCMS (Waters Inc., v4.1) software. Raw data were filtered using the dilution series samples, features which didn't respond linearly to dilution were removed from the dataset. Samples were normalized to median fold change and filtered once again using coefficient of variation in the QC samples; this removed features with low repeatability within the QC samples run throughout the experiment. The remaining features were produced into a table and imported into SIMCA-P 14 software (Umetrics, Umea, Sweden). The most relevant features or metabolites in pair-wise models from MS data were identified using S plots; the 10 most significant features, which cause the separation in the model, were taken forward for identification.

Targeted bile acid MS data were integrated using MassLynx 4.1 and TargetLynx software (Waters Inc. v4.1). Raw Bile Acid MS data (negative ionisation mode only) were imported into TargetLynx and an in house method was used to integrate the bile acid peaks. There were 57 different bile acids integrated using this method and was optimised for this experiment using bile acid standards mixes, which were run at the end of the experiment. Integrated values for each bile acid were exported into an excel spreadsheet where uni-variate analysis was performed (including student's T-test).

5.3.9.3. ^{1}H NMR and UPLC-MS metabolic data analysis

Please refer to Chapter 3, section 3.3.7.3 for details.

5.3.9.4. 16S rRNA gene data analysis

The 16S rRNA sequencing results were pre-processed and analysed using the Mothur program (Schloss et al., 2009) and analysed using the program STAMP. More details of this method have been provided in the Methods section (Chapter 2).

5.4. RESULTS

5.4.1. ^{1}H NMR SPECTROSCOPY DATA PROCESSING

5.4.1.1. Urine metabolic profiling data processing

From the raw data, multivariate statistical models showed that the primary metabolite causing a separation between term and preterm infants in the urinary ^{1}H NMR profiles was creatinine; this metabolite has been shown to be correlated with muscle mass in humans (Baxmann et al., 2008). Babies that are born preterm usually have a lower birth weight and generally have less muscle mass. Since it was evident that this was overshadowing any other difference in the metabolic profile (Figure 5-2), subsequently, the spectral regions corresponding to creatinine (δ3.04, δ4.05) were removed and the dataset was remodelled (Figure 5-3).

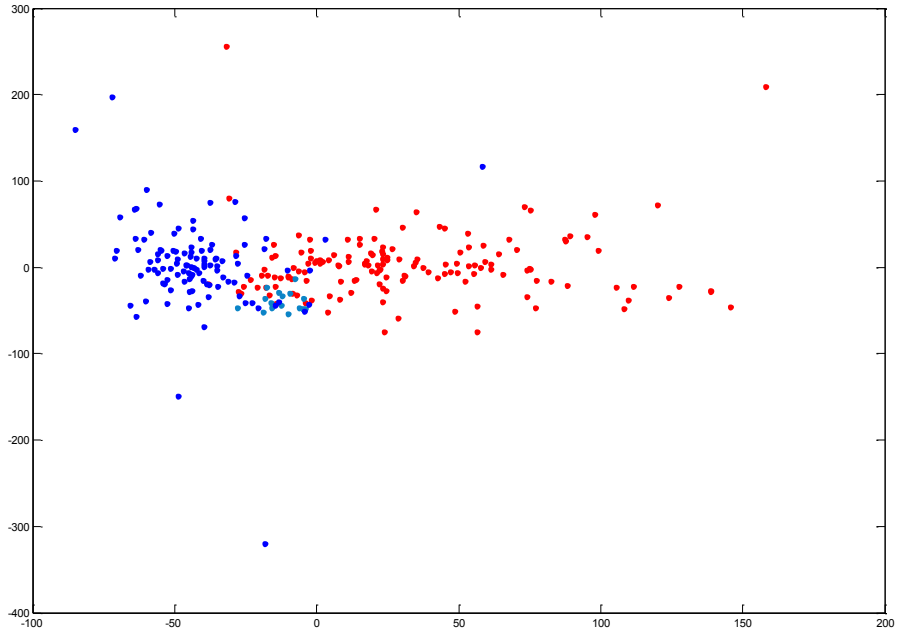


Figure 5-2: PCA model scores plot of global urinary metabolic profiles for each group (Term = blue versus Preterm = red) before the removal of the resonances for the metabolite creatinine (δ3.05 -CH_3; δ4.06 -CH_2), there is a clear separation between groups due to creatinine. Two predictive components, R^2X = 0.25, Q^2=0.221

5.4.1.2. Faecal water metabolic profiling data processing

A proportion (28%) of samples collected for this study were comprised of less than 100mg of stool. From the method development (chapter 3) carried out for this project, the optimum amount of stool needed to conduct metabolic profiling was 100mg but a further the minimum amount of 50mg could be used if necessary. Samples which had a mass of less than 50mg were excluded from analysis (n=57) and samples which were less than 100mg but over 50mg were extracted and acquired. To compensate for the decrease in signal to noise in more dilute samples, the number of scans used to acquire NMR spectral profiles were increased compared with normal acquisition protocols. Samples with 75mg of stool were analysed using double the standard number of scans than 100g samples, and samples with 50mg were analysed with a four-fold increase in the number of scans.

A principal component analysis (PCA) was performed to evaluate whether the number of data acquisition scans had an influence on the variation in the dataset (supplementary Figure S4). Samples with 75mg and 50mg of faecal matter that were analysed using 128 and 256 scans respectively, are interspersed with 100mg samples analysed using the standard 64 scans, indicating no technical bias of this approach after normalisation and no separation between the different groups in the model.

5.4.2. METABOLIC PROFILING OF INFANT URINE AND FAECAL SAMPLES USING ^{1}H NMR SPECTROSCOPY

5.4.2.1. Differences in metabolic signatures of biofluids of infants born term versus preterm

After the removal of creatinine from the urine raw data, principal component analysis (PCA) was used to remodel the urine sample data (Figure 5-3), which demonstrated a clear separation of samples according to their class; born term or preterm. This model includes all time points that contributed to the three component model, explaining 25% of the total variance within the data.

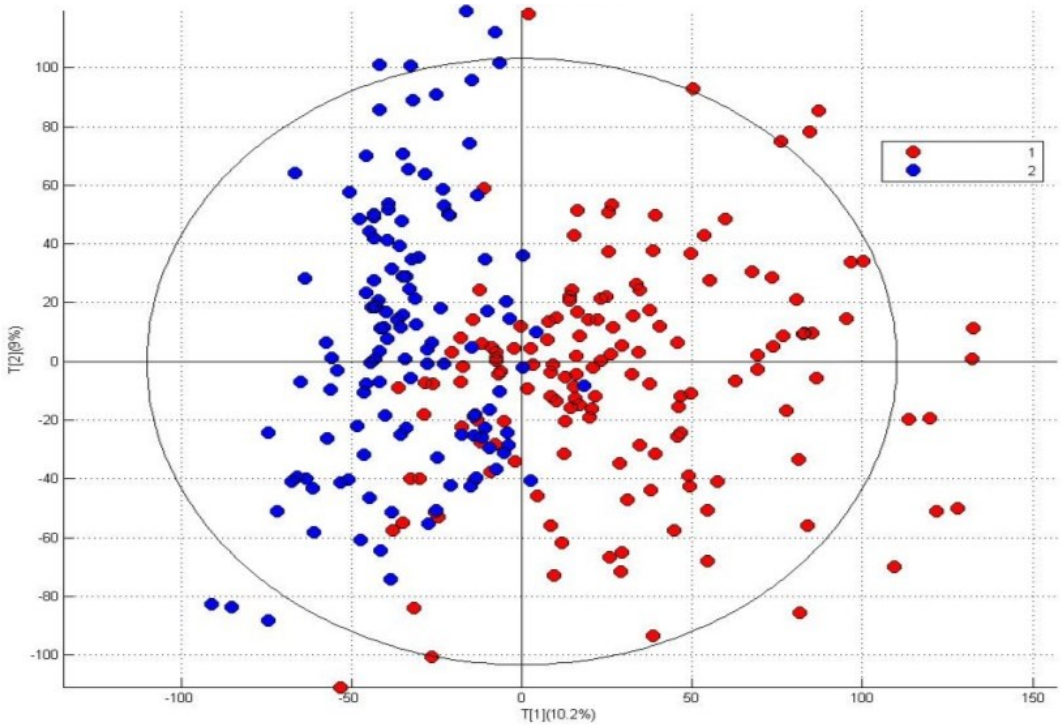


Figure 5-3: Global urinary metabolic profiles for each group, Term =blue, Preterm = red using 1D ^{1}H NMR spectral data. PCA model scores plot, with three predictive components, R^2X = 0.25 and Q^2X=0.22.

A supervised model (OPLS-DA) was built with two orthogonal components resulting in a R^2Y value of 85% and a Q^2Y value of 82% from the urine data; it was looking at the difference between term and preterm infant samples. For this model the metabolites with the highest correlation coefficient (Figure 5-4) were identified and the assignments have been summarised in table 5-1. In particular, a specific metabolite signature including betaine (δ3.27, δ3.9) as well as dimethylamine (δ2.72) and dimethylglycine (δ2.93) was positively correlated to infants born term as well as trimethylamine (TMA) (δ2.91). Term infants were also associated with pseudouridine (δ7.68) and malonate (δ3.12).

The infants born preterm are positively correlated to N-1-methylnicotinamide (NMN) (δ4.48, δ8.19, δ8.91, δ8.97, δ9.27) and myo-inositol (δ3.53) in their urine as well as lysine (δ1.73). Generally, preterm infant's urine also has a higher amounts of sugars within their urine, including mannitol (δ3.69), fructose (δ3.82) as well as unidentifiable sugars (δ3.84, δ4.14, δ3.78).

Table 5-2 – List of urinary metabolites assigned in the loadings plot of OPLS-DA models (shown in Figure 4)

Metabolite name	**^{1}H NMR δ (multiplicity group)**	**Group observed**
Dimethylamine	2.72(s)	Term
Dimethylglycine	2.93(s)	Term
Betaine	3.27(s), 3.9(s)	Term
Pseudouridine	7.68(s)	Term
Malonate	3.12(s)	Term
Trimethylamine (TMA)	2.91(s)	Term
Unknown 1	2.78(s)	Term
1-N-methylnicotinamide	4.48 (s), 8.19(t), 8.91 (d), 8.97(d), 9.27(d)	Preterm
Lysine	1.73 (q)	Preterm
Fructose	3.82 (s)	Preterm
Myo-inositol	3.53 (s)	Preterm
Mannitol	3.69 (s)	Preterm

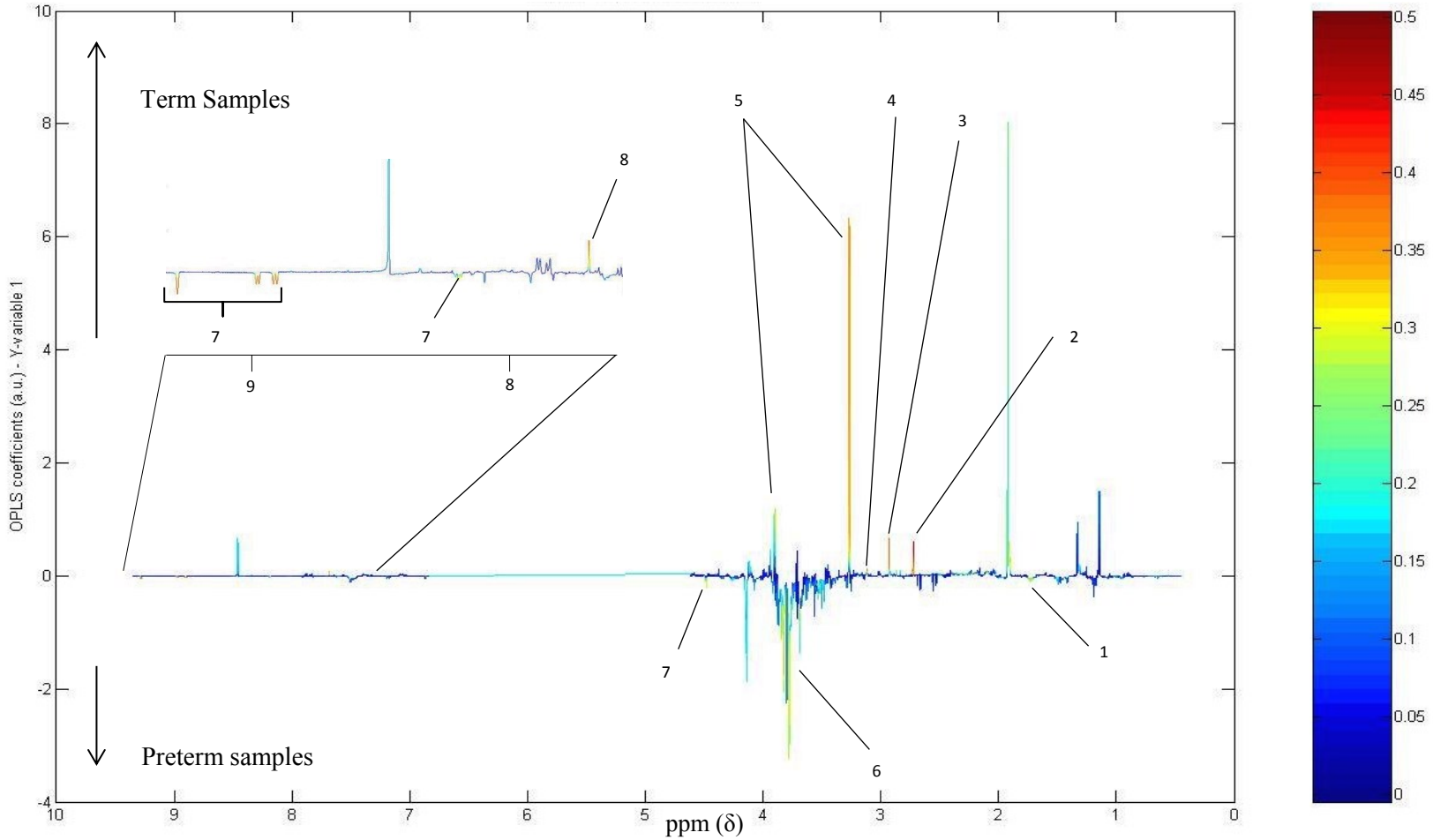


Figure 5-4: Correlation Coefficient plot from OPLS-DA model (R^2X= 0.13, R^2Y=0.85 and Q^2Y=0.82) showing significant metabolites in urine separating between infants born term (positive) versus infants born preterm (negative). Metabolite ID: 1. Lysine (δ1.73), 2. Dimethylamine (δ2.72) 3. Dimethylglycine (δ2.93), 4. Malonate (δ3.12), 5. Betaine (δ3.27, δ3.9), 6. Sugar (δ3.84), Fructose (δ3.82), Mannitol (δ3.69) and Myo-inositol (δ3.53) 7. 1-*N*-methylnicotinamide (δ4.48, δ8.19, δ8.91, δ8.97, δ9.27), 8. Pseudouridine (δ7.68).

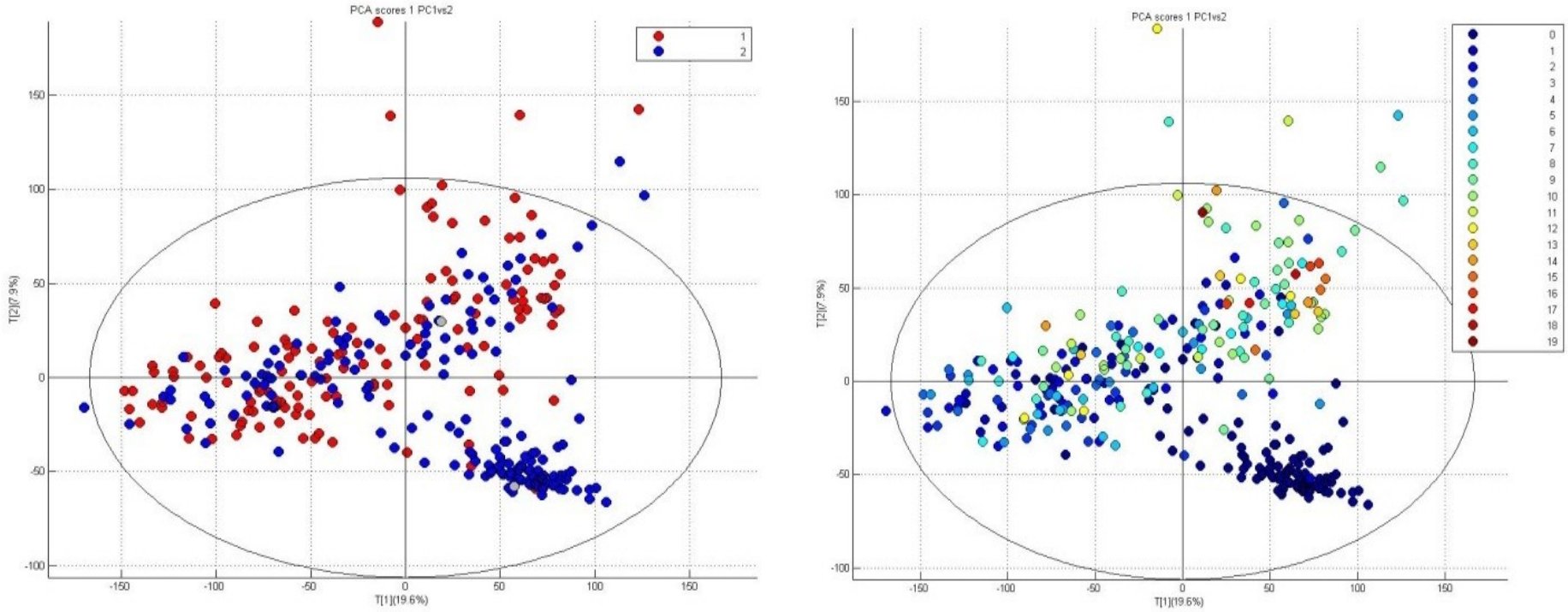


Figure 5-5: PCA scores plot of global faecal water metabolic profile using 1D ^{1}H NMR spectral data, R^2X= 0.27, Q^2X= 0.262 a) scores plot coloured blue = term samples, red = preterm samples. b) scores plot coloured by postpartum age that samples were taken. Colours = Birth sample (blue) through to Week 19 sample (red).

Unsupervised statistical modelling (PCA) was performed on faecal samples (Figure 5-5a); there was some separation of samples according to class (term *vs.* preterm), but much less so compared with urine samples run by the same analysis. By colouring the samples in the PCA scores plot in the same model by sample time point (Figure 5-5b), it can be seen that the separation is dominated by samples collected in the first week postpartum compared to those collected at a later stage. From this plot, there is a clear indication of metabolic profile similarity at birth, as these samples cluster together, but as the infant ages the metabolic profile becomes more variable and distinctive compared to these birth samples.

Table 5-3 – List of faecal water metabolites assigned in the loadings plot of OPLS-DA models (shown in Figure 6)

Metabolite name	**^{1}H NMR δ (multiplicity group)**	**Group observed**
Bile acids	0.73 (s)	Term
Cholate	1.12 (s)	Term
Thiamine	2.57 (s)	Preterm
Glycerol	3.66 (m)	Preterm

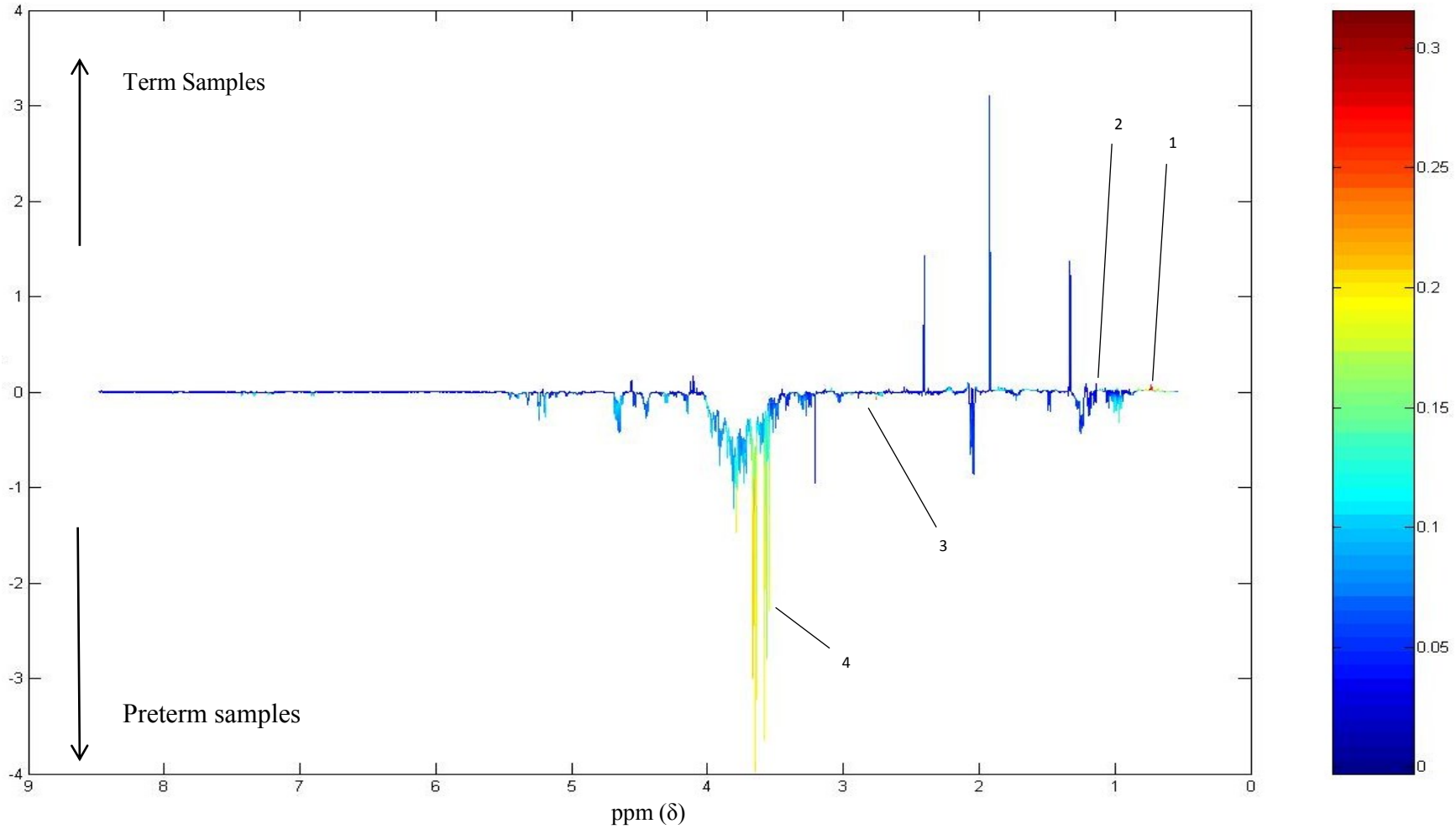


Figure 5-6: Correlation Coefficient plot from OPLS-DA model (R^2X= 0.26, and Q^2Y=0.5) showing significant metabolites in faecal water separating between infants born term (positive) versus infants born preterm (negative). Metabolite ID: 1. Bile acids (δ0.73), 2. Cholate (δ1.12), 3. Thiamine (δ2.57), 4. Glycerol (δ3.66).

A supervised model (O-PLS-DA) was built with two orthogonal components resulting in a R^2X value of 26% and a Q^2Y value of 50% on the faecal data when looking at the difference between term and preterm infant samples; this model is not as strong as the urine spectral data when classifying samples by these groups. The correlation coefficient loadings plot (Figure 5-6) shows the discriminatory metabolites associated with the two classifying group samples, which are summarised in table 5-2. From the plot, there is a higher correlation of bile acids (δ0.73) and cholate (δ1.12) in the term born infant samples compared to the preterm born babies. The loadings plot shows that preterm infants have a higher concentration of glycerol (δ3.66) compared to term infants. Furthermore, preterm infants' faecal water metabolic profile is related to thiamine (δ2.57), also known as vitamin B1.

5.4.2.2. Differences in metabolic signatures from infants born term versus preterm at birth

Preterm infants are born at a gestational age lower than 37 weeks and much of their development happens outside of the womb; this includes organ development. Therefore, the baseline urinary metabolic profile at birth from infants born at term versus preterm is very interesting to investigate further. However, due to the nature of infants born preterm, they are typically born smaller and may have a number of different disorders, thus their first substantial urine samples that could be acquired by NMR was typically obtained a few days from birth rather that immediately. Therefore, for this model preterm infant's urine, which was collected within two weeks from birth, is regarded as a birth sample and compared to infants born at term where their samples were taken at birth (typically within the first 24 hours of life).

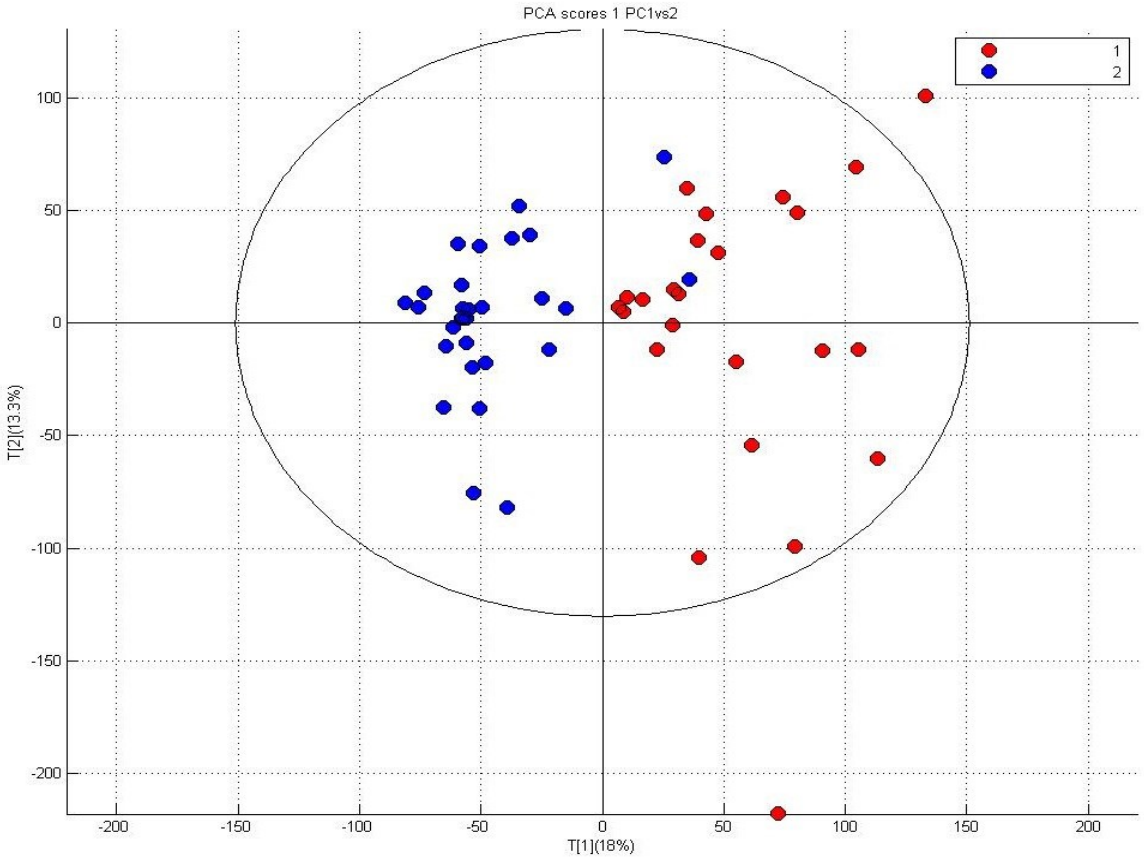


Figure 5-7: PCA of 1D ^{1}H NMR spectra urinary spectral data from infants born term vs preterm taken at birth: blue = term samples, red = preterm samples. R^2X = 0.31 and Q^2X=0.2.

A PCA model was produced using the urine samples (Figure 5-7), which demonstrated a clear separation of samples according to their class; term (n=29) or preterm samples (n=26) taken at birth. From the plot, there are two term samples which are clustered with the preterm samples, these samples are from two infants born at 38 and 39 weeks of gestation. Comparisons of ^{1}H NMR urinary spectra between term and preterm infants at birth was also performed using the supervised method; OPLS-DA. Results from this model are summarised in table 5-3 and the correlation coefficient (Figure 5-8) shows the metabolites that cause the separation of these two classes.

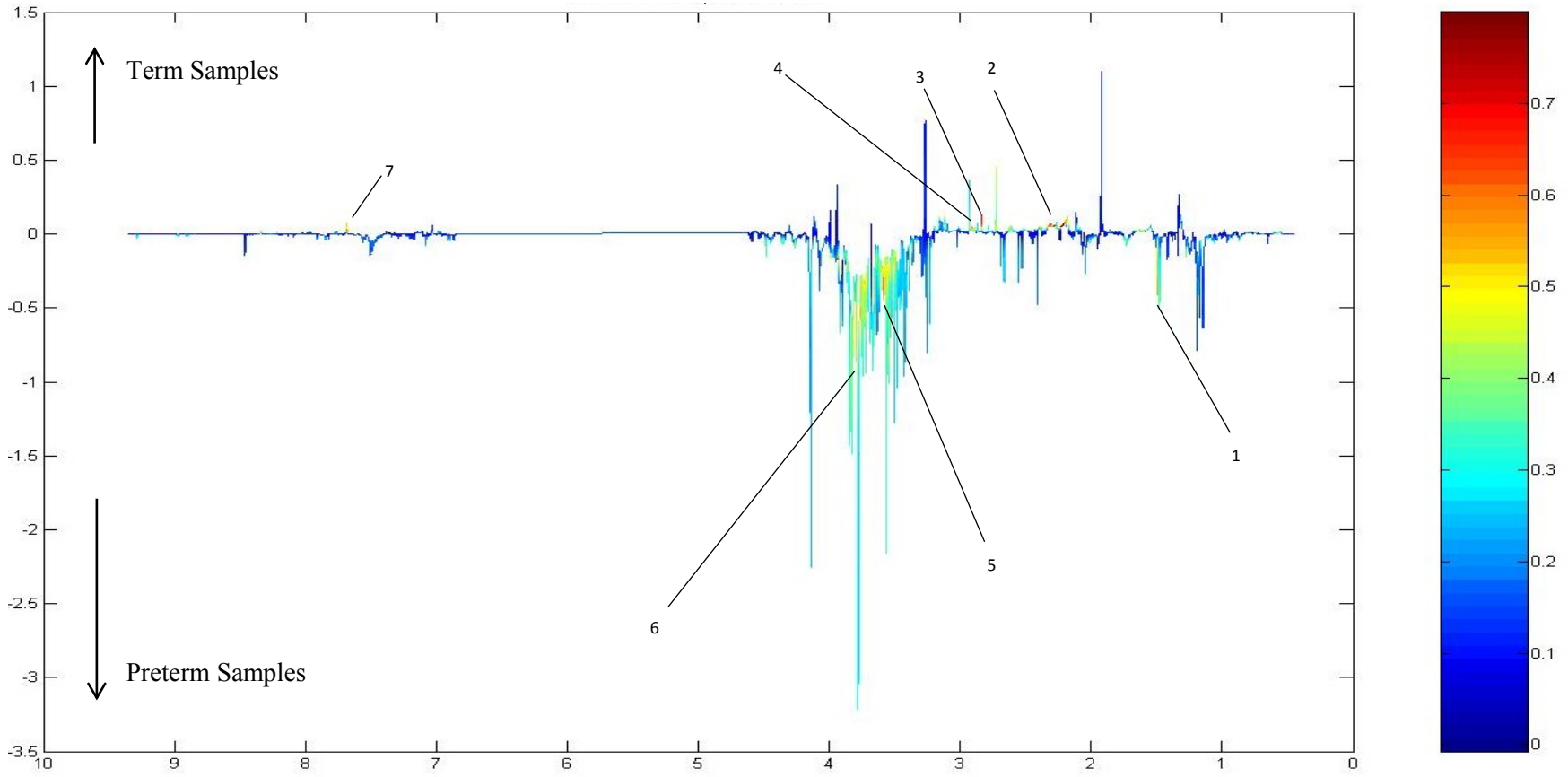


Figure 5-8: Correlation Coefficient plot from OPLS-DA model (R^2X= 0.21, and Q^2Y=0.06) showing significant metabolites in urine separating between infants born term (positive) versus infants born preterm (negative) taken at birth. Metabolite ID: 1. Alpha-aminoisobutyrate (δ1.49), 2. Unknown (δ2.31), 3. Methylguanine (δ2.83), 4. Trimethylamine (TMA) (δ2.91), 5. Threonine (δ3.59), 6. Myo-inositol (δ3.65), 7. Pseudouridine (δ7.68).

Similarly to the results from the previous model, urine samples taken at birth from term born infants were positively correlated with methylguanidine (δ2.83), trimethylamine (TMA) (δ2.905), and pseudouridine (δ7.68) as well as an unknown metabolite (δ2.31). This may suggest that birth samples are again dissimilar from samples taken after a few days of life. Urine from preterm infants at birth is related to threonine (δ3.59), 1-*N*-methylnicotinamide (δ4.48) and myo-inositiol (δ3.65) as well as an unknown metabolite, which was tentatively assigned as α-aminoisobutyrate based on its chemical shift (δ 1.49) and signal multiplicity.

Table 5-4 – List of urinary metabolites assigned in the loadings plot of OPLS-DA models (shown in Figure 5-8)

Metabolite name	**^{1}H NMR δ (multiplicity group)**	**Group observed**
Pseudouridine	7.68 (s)	Term
Methylguanidine	2.83 (s)	Term
Trimethylamine	2.91 (d)	Term
Unknown 1	2.31 (s)	Term
Alpha-aminoisobutyrate	1.49 (s)	Preterm
Threonine	3.59 (d)	Preterm
1-*N*-Methylnicotinamide	4.48 (s)	Preterm
Myo inositol	3.65 (m)	Preterm

A PCA model of faecal samples was calculated for samples taken at birth from infants born term and preterm (Figure 5-9); there was separation according to class and explains 37% of the variability within the data. This model has better predictability compared to the urine model which can be recognised from the Q^2X values; 0. 2 = urine vs faecal = 0.31. In this scores scatter plot, there is a clear indication of metabolic profile similarity for the term infants compared to the preterm infant's samples, which were highly variable in composition.

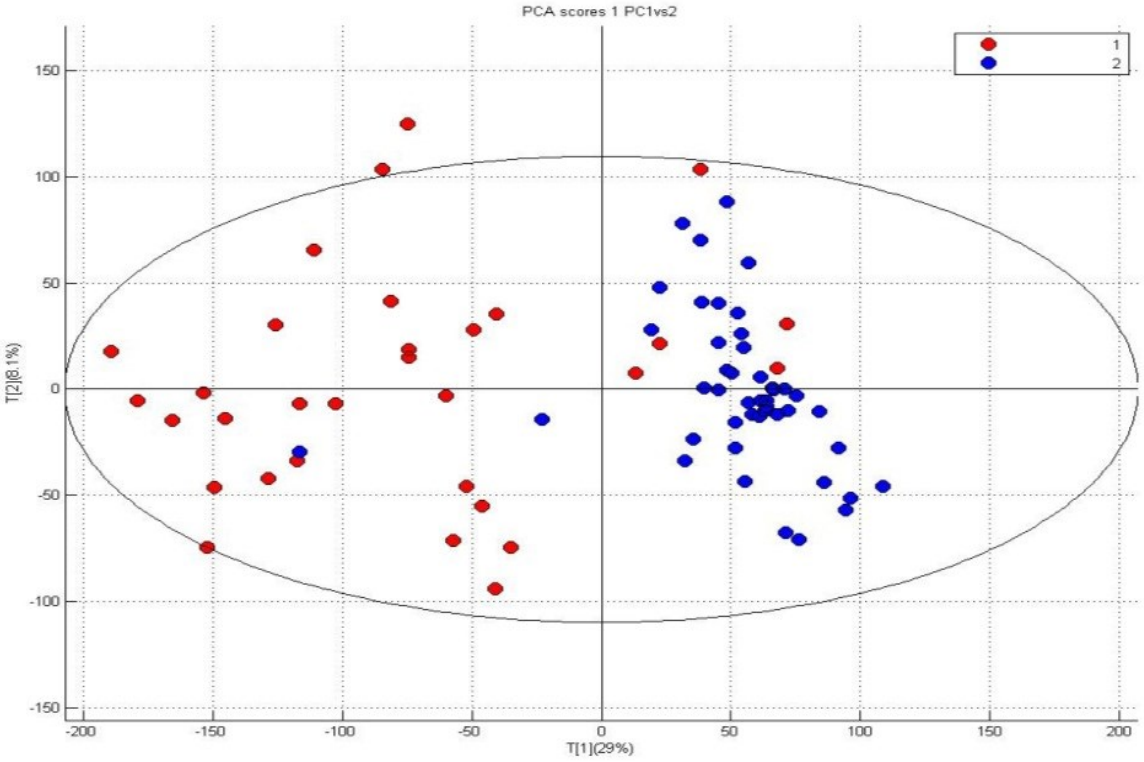


Figure 5-9: PCA of 1D ^{1}H NMR spectra faecal water spectral data from infants born term vs preterm taken at birth: blue = term samples, red = preterm samples. R^2X = 0.37 and Q^2X=0.31.

Table 5-5 – List of faecal water discriminatory metabolites assigned in the correlation coefficient plot of OPLS-DA models (shown in Figure 5-10)		
Metabolite name	**^{1}H NMR δ (multiplicity group)**	**Group observed**
Bile Acid	0.68 (s), 0.73 (s), 0.83 (s)	Term
Unknown	3.09 (t)	Term
Sugars*	3.61 (m), 3.76 (m), 4.14 (d), 4.67 (m), 5.27 (d)	Preterm
Alanine	1.48 (d)	Preterm
Phenylalanine	7.34 (m)	Preterm

The pairwise OPLS-DA model was performed on this stool data to assess the differences between the term and preterm samples taken at birth (figure 5-10); the discriminatory metabolites associated with the two classifying group samples are summarised in table 5-4. From the plot, there was a higher correlation of bile acids (δ0.68, δ0.73, δ0.83) in the term baby samples compared to the preterm babies, which was also seen when modelling the whole data set. Term infants also correlated with an unidentifiable metabolite (δ3.09). When using STOCSY to determine the identity of this metabolite, it correlates to the bile acids at δ0.68, δ0.73 and δ0.83, suggesting that it could be a bile acid conjugated to taurine (Duarte et al., 2009).

Comparatively, there was a higher amount of sugars (δ3.61, δ3.76, δ4.14, δ4.67, δ5.27) in preterm infant's faecal samples compared to term infants. Further to this, preterm infant's faecal samples were related to two amino acids; alanine (δ1.48) and phenylalanine (δ7.34).

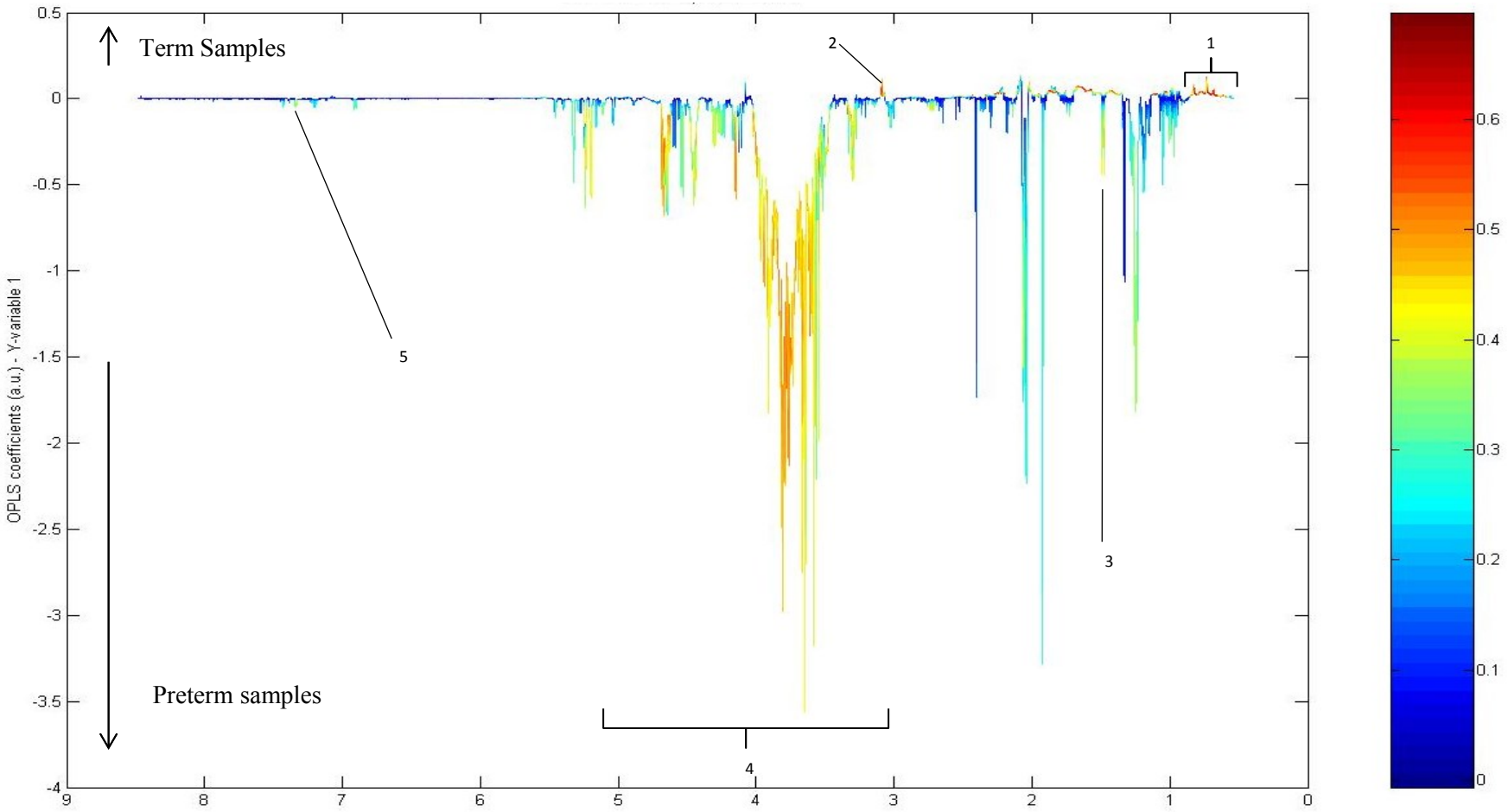


Figure 5-10: Correlation Coefficient plot from OPLS-DA model (R^2X= 0.35, and Q^2Y=0.67) showing significant metabolites in faecal water separating between infants born term (positive) versus infants born preterm (negative) taken at birth. Metabolite ID. 1. Bile acids (δ0.68, δ0.73, δ0.82), 2. Unknown (δ3.09), 3. Alanine (δ1.48), 4. Sugars (δ3.61, δ3.76, δ4.14, δ4.67, δ5.27), 5. Phenylalanine (δ7.34).

5.4.2.3. Differences in metabolic signatures from term birth samples versus preterm samples at term equivalent age

Infants which are born preterm are deemed term equivalent age when they have reached the 37 week of gestation in the weeks spent postpartum. For example, if an infant was born at the 32nd week of gestation, 5 weeks after birth they would be termed a preterm infant at term equivalent age. By comparing term infants birth samples and samples taken from preterm infants when they reach term equivalent age (37-41 gestation week), it gives a clearer idea as to how the preterm infants have developed differently, if at all, compared to infants which are born at term.

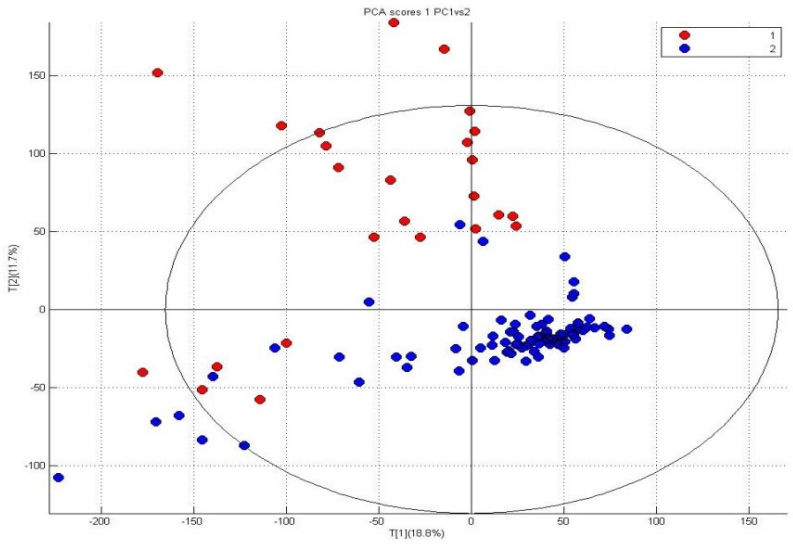


Figure 5-11: PCA of 1D ^{1}H NMR spectra faecal water spectral data from infants born term at birth vs preterm at term equivalent age: blue = term samples, red = preterm samples. R^2X = 0.3 and Q^2X=0.23.

From the urinary NMR data, there was no significant difference in urinary metabolic profiles between term infants' samples taken at birth compared to preterm infant samples taken at term equivalent age (Supplementary Figure S5). However, there was significant separation of stool metabolic profiles according to this class definition (Figure 5-11). In this scores scatter plot, there seemed to be a clear indication of metabolic profile similarity, as term birth samples cluster very closely in this plot whereas preterm samples were highly variable in comparison. When looking at

the variance from this model using the loadings plot (data not shown); in the first principal component, there is a high concentration of sugars in preterm infant faecal samples compared to term samples and in the second principal component, preterm infant samples have higher amounts of alanine (δ1.48) in their faecal samples.

OPLS-DA was performed on the stool data; to further investigate the differences between the term at birth and preterm samples taken at term equivalent age. The correlation coefficient loadings plot (Figure 5-12) shows the discriminatory metabolites associated with the two classifying group samples, which are summarised in table 5-5. From the plot, there is a higher correlation of bile acids (δ0.68, δ0.73, δ0.83) in the term baby samples compared to the preterm babies, which was seen in the previous model when comparing term versus preterm at birth.

Table 5-6 – List of metabolites assigned in the loadings plot of OPLS-DA models (shown in Figure 5-12)

Metabolite name	**^{1}H NMR δ (multiplicity group)**	**Group observed**
Bile Acid	0.68 (s), 0.73 (s), 0.83 (s)	Term
Leucine	0.97 (m)	Preterm
Alanine	1.48 (d)	Preterm
Lysine	3.03 (m)	Preterm
Tyramine	6.92 (m), 2.93 (t)	Preterm

The loadings plot indicates a stronger correlation of amino acids - including alanine (δ1.48), leucine (δ0.97), lysine (δ3.03) - and tyramine (δ6.92, δ2.93) in preterm infant's faecal samples compared to term infants.

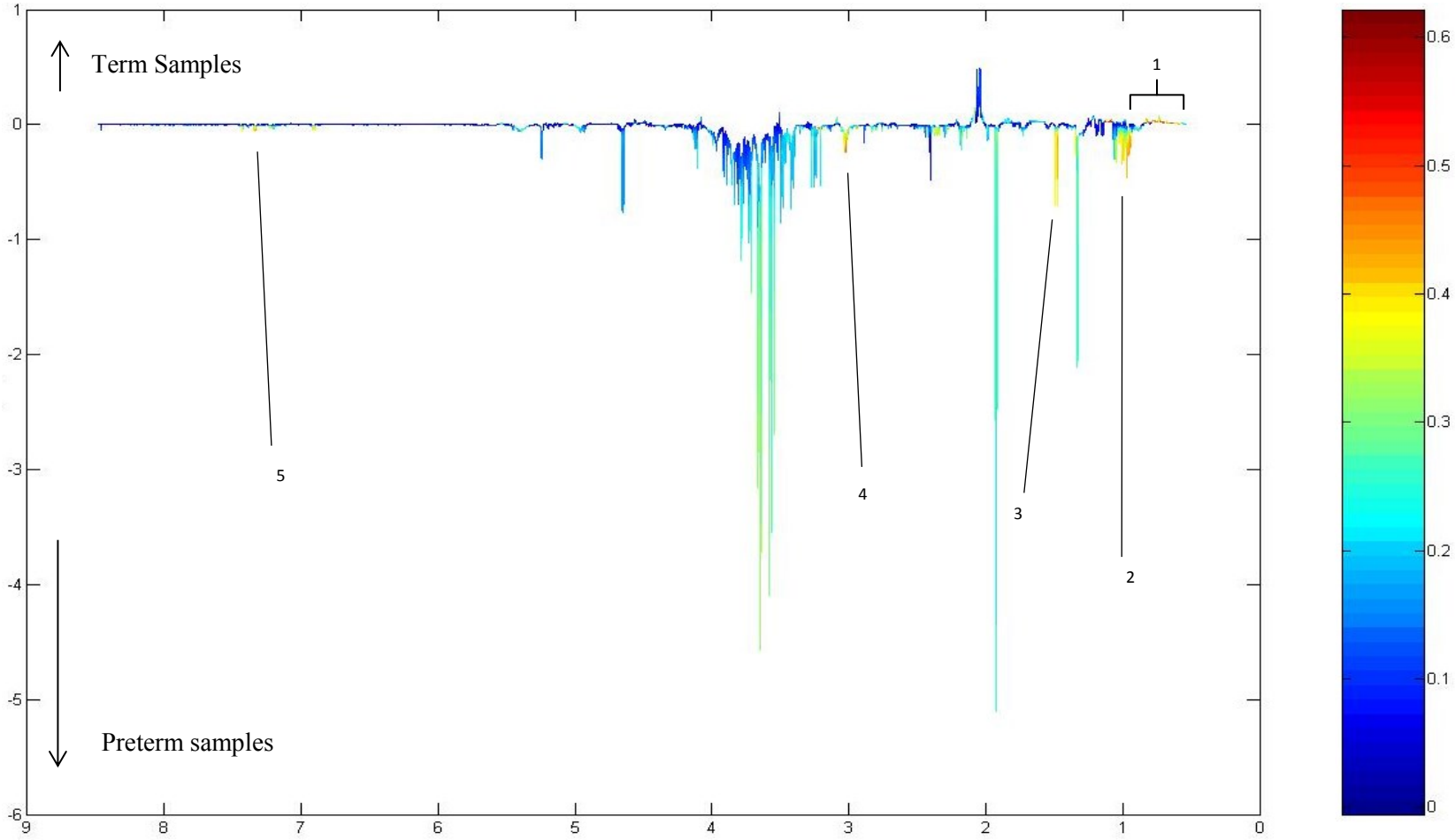


Figure 5-12: Correlation Coefficient plot from OPLS-DA model (R^2X= 0.27, and Q^2Y=0.8) showing significant metabolites in faecal water separating between infants born term taken at birth (positive) versus infants born preterm (negative) taken at term equivalent age. Metabolite ID: 1. Bile acid (δ0.68, δ0.73, δ0.83), 2. Leucine (δ0.97), 3. Alanine (δ1.48), 4. Lysine (δ3.03), 5. Tyramine (δ6.92).

Due to birth or meconium samples being very distinctive compared to faecal samples taken later in life - even a few days later – and causing the most amount of discrimination in the profiles, term samples which were taken at one week postpartum were modelled against preterm samples taken at term equivalent age using PCA (Figure 5-13).

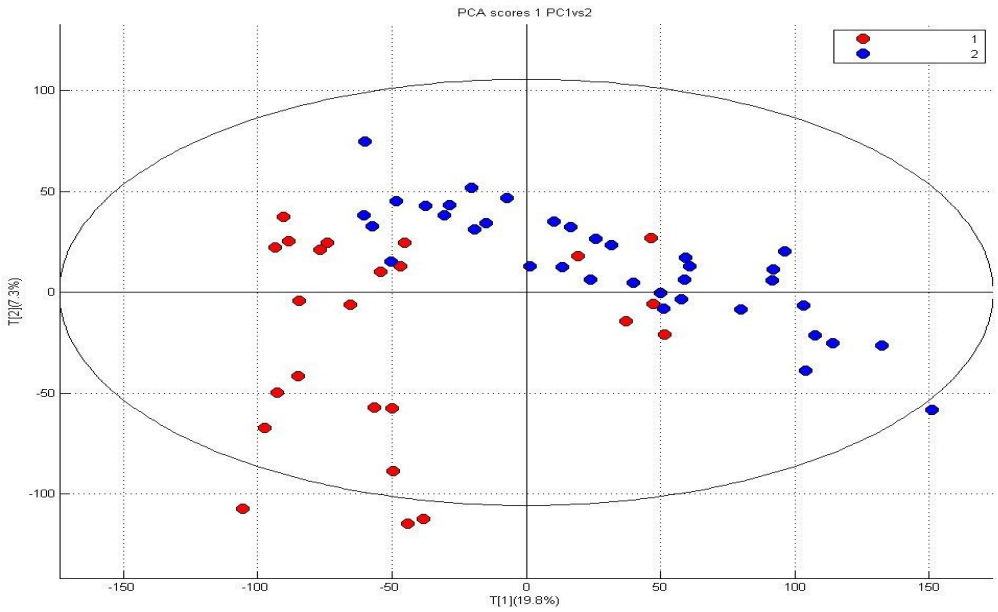


Figure 5-13: PCA of 1D ^{1}H NMR spectra faecal water spectral data from infants born term at one week vs preterm at term equivalent age: blue = term samples, red = preterm samples. R^2X = 0.3 and Q^2X=0.23.

A supervised statistical method was performed (OPLS-DA); the correlation coefficient plot (Figure 5-14) showing the discriminatory metabolites associated with each of the two classes and are further summarised in table 5-6.

Table 5-7 – List of metabolites assigned in the loadings plot of OPLS-DA models (shown in Figure 5-14)

Metabolite name	**^{1}H NMR δ (multiplicity group)**	**Group observed**
Tyramine	3.23 (s), 2.95 (m), 7.23 (d)	Term
N-acetyl-glycoproteins	2.03 (s) 2.06 (s)	Preterm

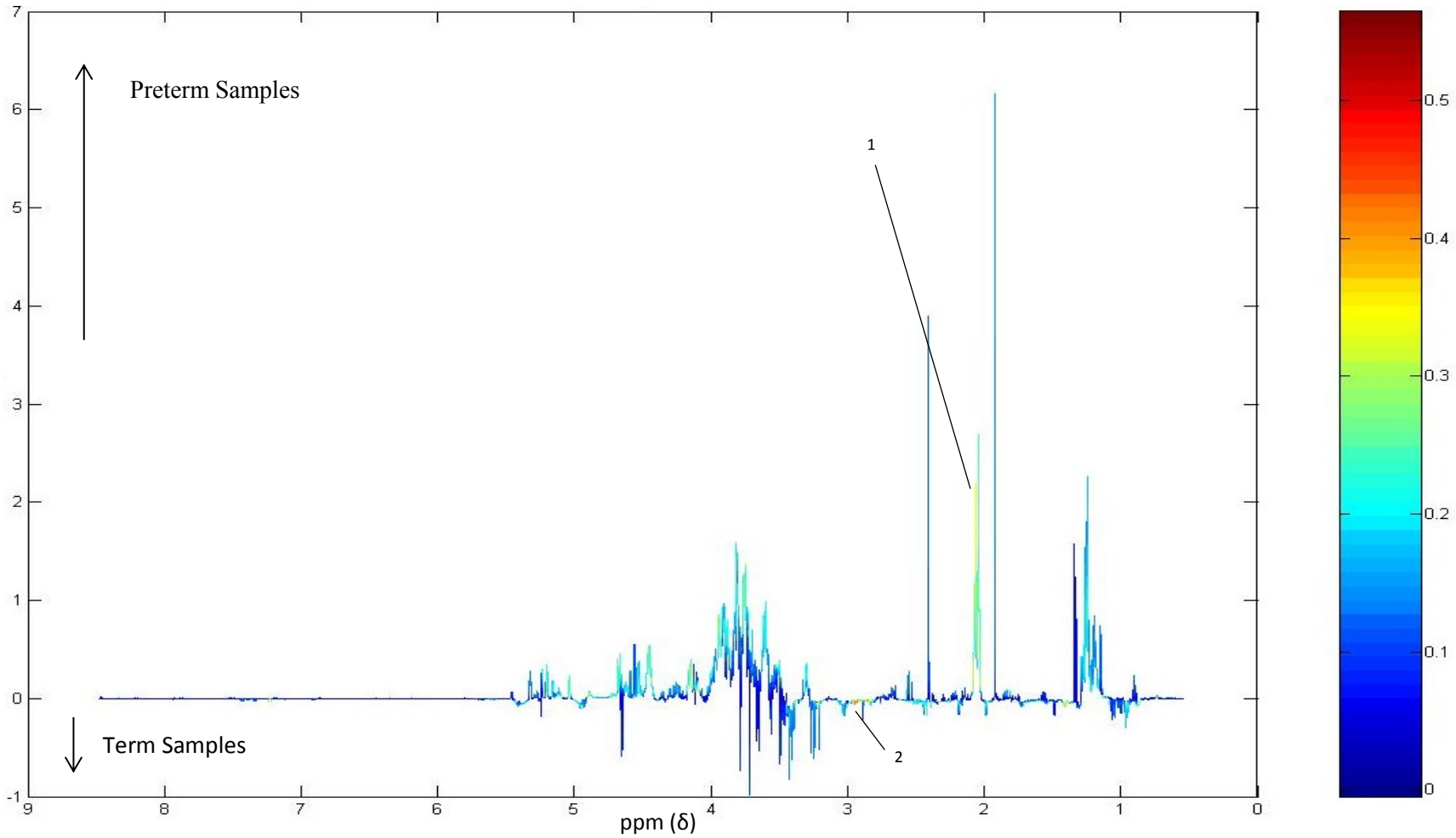


Figure 5-14: Correlation Coefficient plot from OPLS-DA model (R^2X= 0.27, and Q^2Y=0.8) showing significant metabolites in faecal water separating between infants born term taken at one week (negative) versus infants born preterm (positive) taken at term equivalent age. Metabolite ID: 1. N-acetyl-glycoprotein (δ2.03, δ2.06), 2. Tyramine (δ2.95).

In the previous model, term birth samples were modelled against preterm samples, the preterm samples were associated with tyramine (δ3.23, δ2.95, δ7.23). However, in this updated model, term samples were taken at one week postpartum and modelled against preterm samples taken at term equivalent age, tyramine was associated with term infants. Preterm samples taken at term equivalent age were related to *N*-acetyl glycoproteins (δ 2.03, 2.04) when modelling against term faecal samples taken at one week postpartum.

5.4.2.4. Disparities in metabolic profiles from samples taken at 3 months postpartum from infants born term versus preterm.

A principal components analysis of urine samples was performed (Figure 5-15), which demonstrated a clear separation of samples according to their class; term (n=26) or preterm samples (n=17) taken at three months postpartum. OPLS-DA was performed on the urine data to further assess the difference between term and preterm samples taken three months postpartum. For this pairwise OPLS-DA model the metabolites with the highest correlation coefficient (Figure 5-16) were identified and metabolite assignments are summarised in table 5-7.

Table 5-8 – List of urinary metabolites assigned in the loadings plot of OPLS-DA models (shown in Figure 5-16)

Metabolite name	^{1}H NMR δ (multiplicity group)	Group observed
Betaine	3.27 (s)	Term
Pseudouridine	7.68 (s)	Term
Dimethylglycine	2.93 (s)	Term
Dimethylamine	2.72 (s)	Term
N-acetylglutamate	2.05 (m), 2.04	Term
Mannitol	3.82 (s)	Preterm
Bile Acid	0.76 (s)	Preterm
1-*N*-methylnicotinamide	4.48 (s), 8.19 (t), 8.97 (d), 9.28 (s)	Preterm
Fructose	3.78 (m)	Preterm
Lysine	1.73 (q)	Preterm

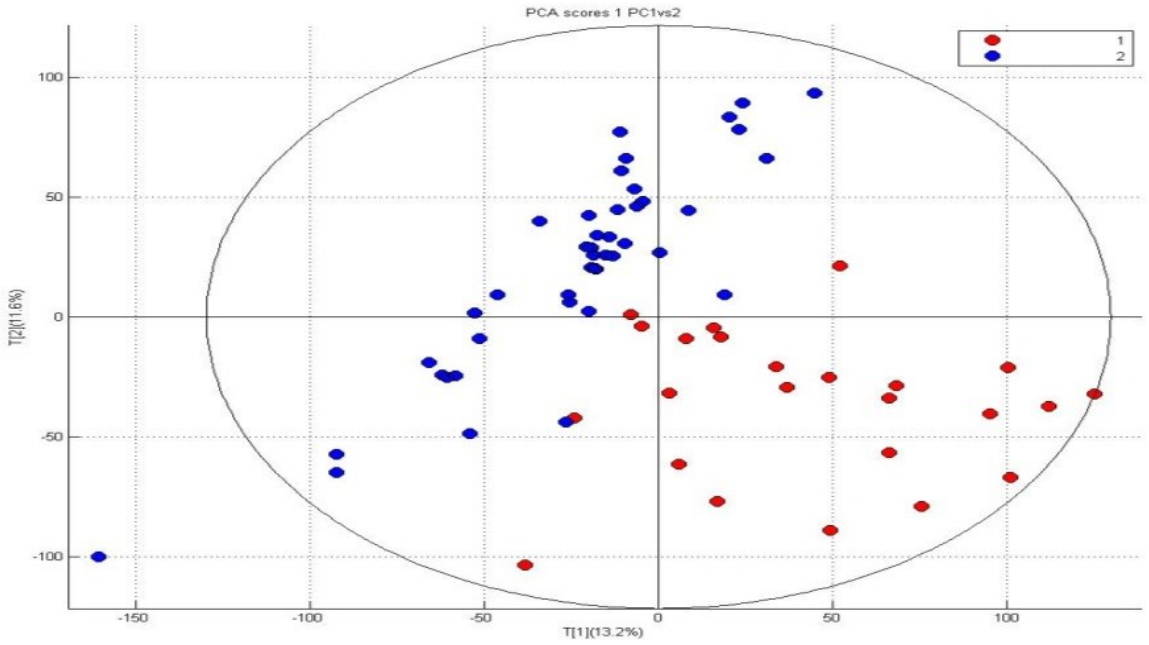


Figure 5-15: PCA of 1D ^{1}H NMR urine spectral data from infants born term vs preterm at 3 months' postpartum: blue = term samples, red = preterm samples. R^2X = 0.377 and Q^2X=0.18.

Term infant's urine at three months postpartum is related to betaine (δ3.27), dimethylglycine (δ2.93), dimethylamine (δ2.72); these metabolites were also discriminatory when modelling all the data points regardless on time point. *N*-acetylglutamate (δ2.04, δ2.05) was also associated with term infants at 3 months postpartum. Preterm infant's urine at three months postpartum had a higher concentration of mannitol (δ3.82), 1-*N*-methylnicotinamide (δ4.48, δ8.19, δ8.97, δ9.28), fructose (δ3.78), lysine (δ1.73) as well as an unknown species of bile acid (δ0.76).

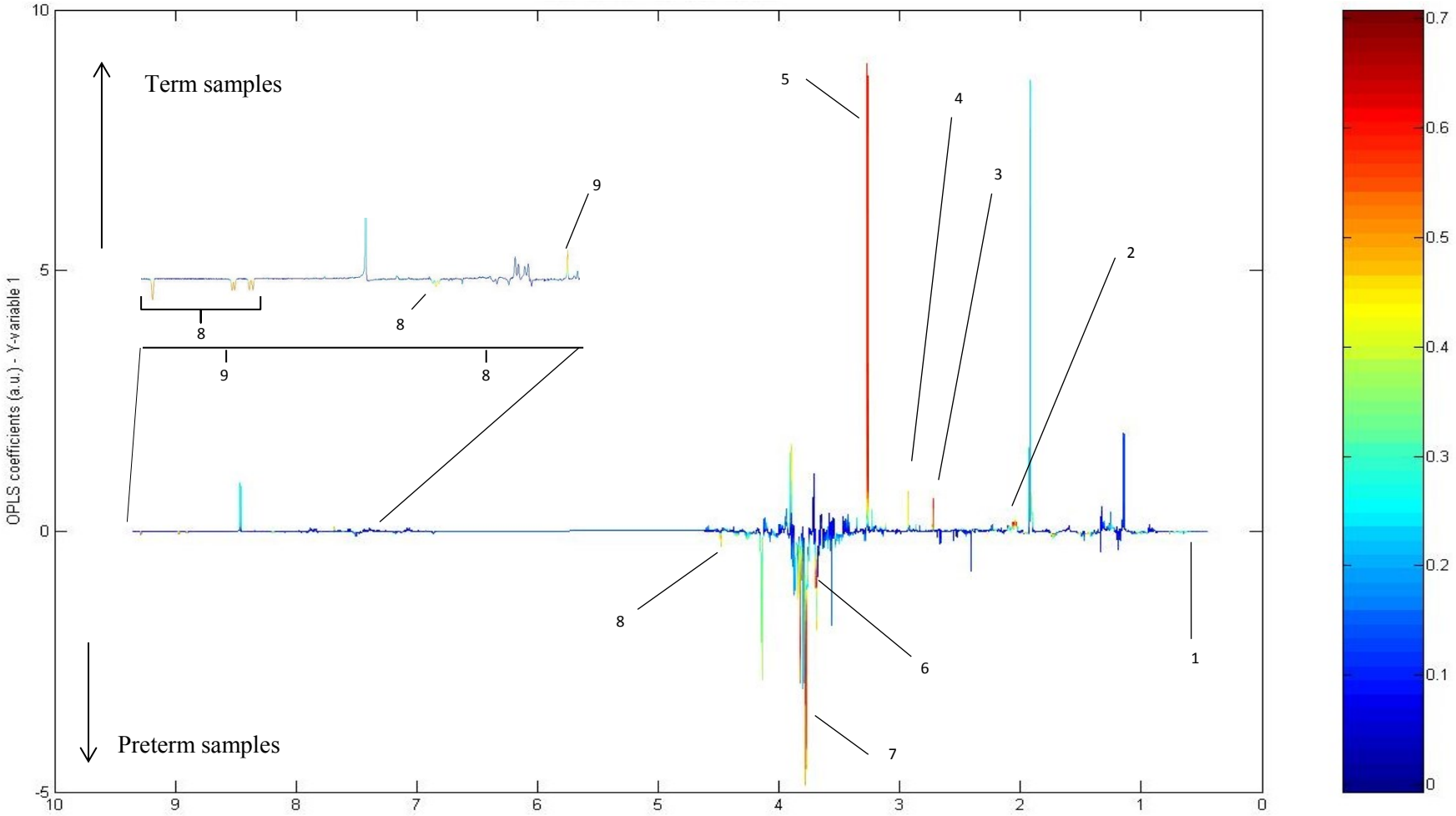


Figure 5-16: Correlation Coefficient plot from OPLS-DA model (R^2X= 0.3, and Q^2Y=0.9) showing significant metabolites in faecal water separating between infants born term (positive) versus infants born preterm (negative) taken at three months postpartum. Metabolite ID: 1. Bile acids (δ0.76), 2. N-acetyl-glycoproteins (δ2.04, δ2.06), 3. Dimethylamine (δ2.73), 4. Dimethylglycine (δ2.93), 5. Betaine (δ3.27), 6. Fructose (δ3.78), 7. Mannitol (δ3.82), 8. 1-N-methylnicotinamide (δ4.48, δ8.19, δ8.97, δ9.28), 9. Pseudouridine (δ7.68).

These infants were on different types of nutrition, which could contribute to their development and thus their metabolic profile; including being fed with breast milk, formula milk, parenteral nutrition or a combination of parenteral nutrition and formula milk. By re-colouring the scores plot of the previous PCA model by the type of feeding group the infants were on (Figure 5-17), it is clear that within the term born infants, there was no clear separation between infants which were fed breast milk versus formula fed in this model. But from the scores plot, the largest difference was seen between preterm infants, which were fed parenteral nutrition only compared to those which had some formula feeding (combination feeding) making them more like their term born counterparts. One sample, which was labelled as deriving from a breast milk fed baby clearly resembled those in metabolic composition from parenteral nutrition; this sample was most likely labelled incorrectly.

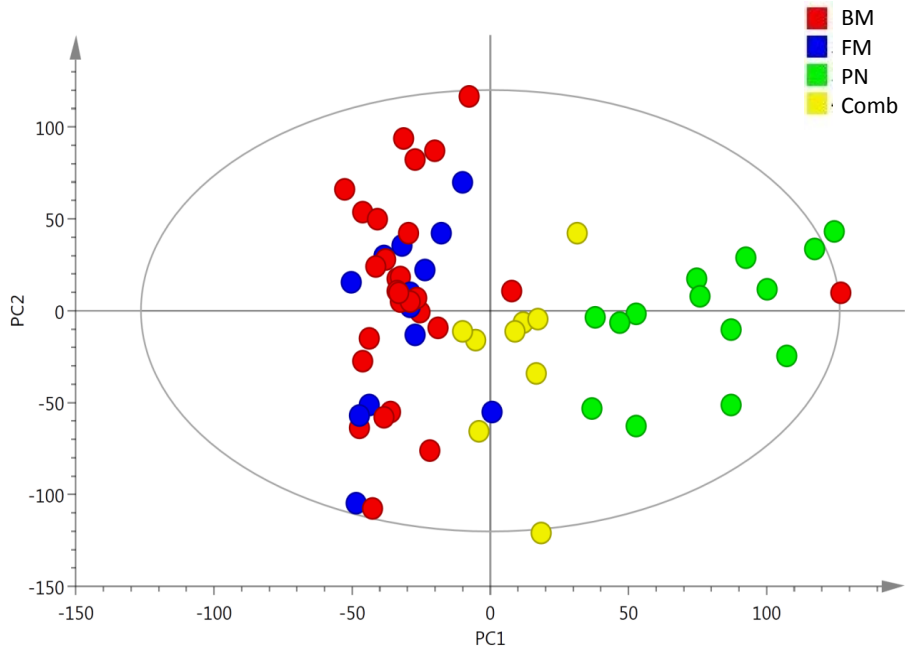


Figure 5-17: PCA of 1D ^{1}H NMR urine spectral data from infants born term vs preterm at 3 months postpartum:, red = Breast milk, blue = Formula milk, Green = parenteral nutrition, Yellow = combination of parenteral nutrition and formula/breast milk. R^2X = 0.377 and Q^2X=0.18.

Conversely, faecal water metabolic profiles at three months' postpartum age show no separation between infants born at term compared to preterm born infants. This suggests that infant's faecal water metabolic profiles are very similar once they are at this age. This may be due to the same amount of time outside the womb and therefore exposure to the environment is more influential than human development on the faecal metabolic profiling (supplementary Figure S6).

5.4.3. GLOBAL URINARY METABOLIC PROFILING HILIC-MS

5.4.3.1. Metabolic profiling using HILIC-MS of infants born term vs. preterm at all time points

A principal component model performed on urine samples (Figure 5-18), demonstrated a clear separation of samples according to their class; born term (n=12) or preterm (n=55). Principal component one shows a clear separation between infants born term vs preterm, whereas the variation represented in component two is from the preterm infants' samples increasing with postpartum age (Supplementary figure S7). Samples from term infants were only obtained at birth, whereas samples form preterm infants were gained from birth until 19 weeks postpartum. The supervised method of OPLS-DA was performed by separating samples into two classes, for both negative and positive ionisation mode of the urine MS data. Looking at the features, which cause the separation in urine data between the two groups, a number of metabolites were identified and are summarised in table 5-8.

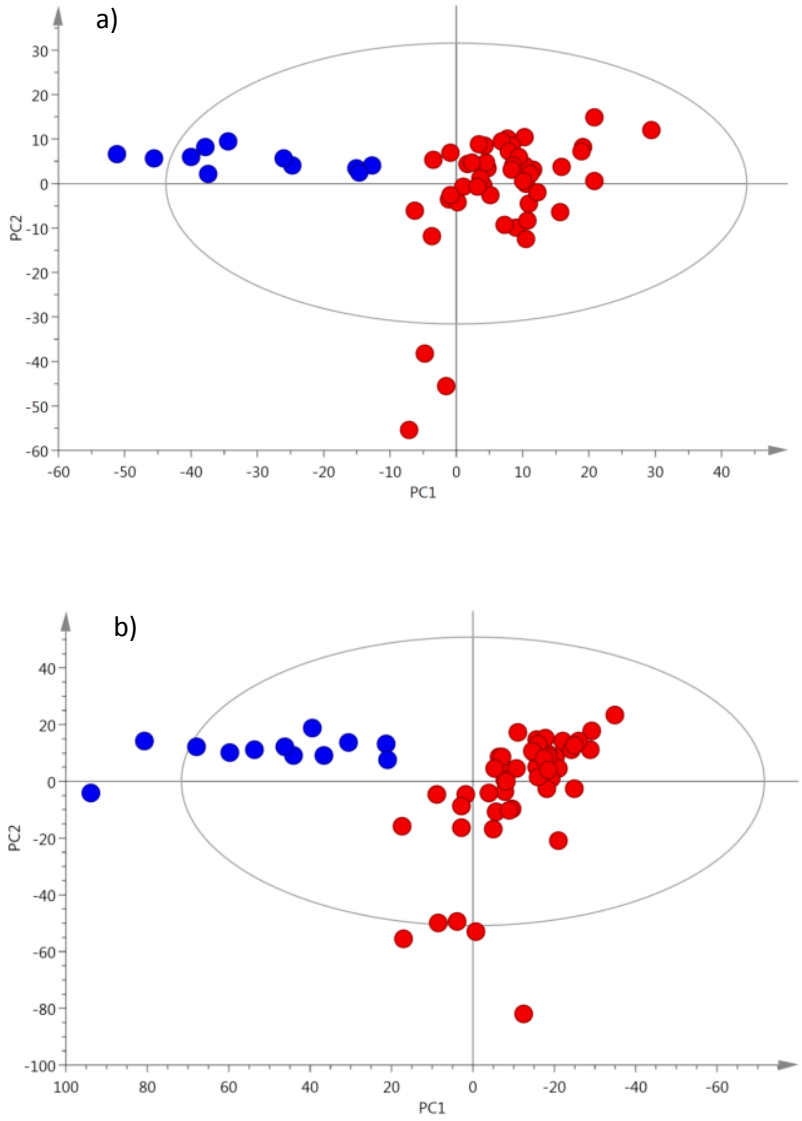


Figure 5-18: PCA of Global HILIC MS urine data from infants born term vs preterm; red = preterm, blue =term. Both models were built with four predictive components a) negative ionization mode $R^2X = 0.327$, $Q^2X = 0.097$. b) positive ionization mode $R^2X = 0.327$, $Q^2X = 0.1$

For the negative ionisation mode data, infants who were born term separated due to metabolites including hydroylprolyl-leucine, a four polypeptide chain and acetylhomoserine. For the positive ionisation mode, infants who were born term separate due to metabolites including homocarnosine, linoleoylglycerophosphocholine and phenylalanyl-threonine.

For the negative ionisation mode data, infants which were born preterm separated due to relatively higher concentrations of histidine and ascorbic acid, as well as *N*-acetylgalactosamine 6–sulphate, Indolepyruvate and hydroxyproline. For the positive ionisation mode data, infants which were born preterm separated due to metabolites including 7-methylguanine and 6-hydroxynicotinic acid.

Table 5-9 - List of features assigned from urinary HILIC-MS data from S-plot of the OPLS-DA models between term and preterm.			
Potential Metabolite ID	**Mz/RT**	**Group Observed**	**Level of assignment**
Homogentistic acid	167/80	Term (ESI –ve)	2
Hydroylprolyl-Leucine	225.1/301	Term (ESI –ve)	2
Asp Ile Val Tyr	507.2/301	Term (ESI –ve)	3
Acetylhomoserine	160.1/345	Term (ESI –ve)	2
Unknown 1	511.3/347	Term (ESI –ve)	4
N-acetylgalactosamine-6-sulphate	301/136	Preterm (ESI –ve)	2
Indolepyruvate	202/349	Preterm (ESI –ve)	2
Hydroxyproline	130.1/364	Preterm (ESI –ve)	2+
Histidine	154.1/525	Preterm (ESI –ve)	2+
Ascorbic Acid	175/567	Preterm (ESI –ve)	2
Phenylalanyl Threonine	267.1/302	Term (ESI +ve)	2
Linoleoylglycerophosphocholine	535.3/349	Term (ESI +ve)	2
Unknown 2	201.2/356	Term (ESI +ve)	4
Unknown 3	377.3/414	Term (ESI +ve)	4
Homocarnosine	242.2/450	Term (ESI +ve)	2
Unknown 4	160.2/569	Term (ESI +ve)	4
7-methylguanine	166.1/171	Preterm (ESI +ve)	2+
6-Hydroxynicotinic Acid	140/559	Preterm (ESI +ve)	2

5.4.3.2. Metabolic profiling using HILIC-MS of infants born term vs. preterm at birth

A PCA model derived from urine samples collected at birth from infants born term or preterm was (Figure 5-19), demonstrated a clear separation of samples according to their class; born term (n=12) or preterm (n=14) at birth. An OPLS-DA was performed by separating samples into two classes, for both negative and positive ionisation mode of the urine data. The results of this model, yielded features which were mainly responsible for the separation in urine data between the two groups a number of metabolites which were identified and summarised in table 5-9.

For the positive ionisation mode, infants who were born term separated due to metabolites including dipeptides and amino acids as well as citrate and homocarnosine. For the positive ionisation mode data, infants who were born preterm separated due to metabolites including glutamine, and 6-hydroxynicotinic acid.

For the negative ionisation mode, infants who were born term separate due to metabolites including hydroylprolyl-leucine, a four polypeptide chain and acetylhomoserine. For the negative ionisation mode data, infants which were born preterm separated due to relatively higher concentrations of ascorbic acid and 1,6-Naphthalenedisulfonic acid.

Table 5-10 - List of variables assigned from urinary HILIC-MS data from S-plot of the OPLS-DA models between term and preterm at birth.

Potential Metabolite ID	Mz/rt	Group Observed	Level of assignment
Homogentistic acid	167/80	Term (ESI –ve)	2
Hydroylprolyl-Leucine	225.1/301	Term (ESI –ve)	2
Asp Ile Val Tyr	507.2/301	Term (ESI –ve)	3
Acetylhomoserine	160.1/345	Term (ESI –ve)	2
1,6-Naphthalenedisulfonic acid	287/285	Preterm (ESI –ve)	2
Ascorbic Acid (Vit C)	175/567	Preterm (ESI –ve)	2
Citrate	193/189	Term (ESI +ve)	2
Phenylalanyl-Threonine	267.1/302	Term (ESI +ve)	3
Tryptophyl-Isoleucine	318.2/411	Term (ESI +ve)	3
Homocarnosine	242.2/450	Term (ESI +ve)	2
Proline	116.1/514	Term (ESI +ve)	2
Glutamine	147/285	Preterm (ESI +ve)	2+
6-Hydroxynicotinic acid	140/559	Preterm (ESI +ve)	2

5.4.3.3. Metabolic profiling using HILIC-MS of Term at birth vs. Preterm at term equivalent age

A principal component analysis of urine samples was performed (Figure 5-20), which demonstrated a clear separation of samples according to their class; born term at birth (n=12) or preterm at term equivalent age (n=19). The supervised method of OPLS-DA was performed by separating samples into two classes, on both negative and positive ionisation mode of the urine data. Looking at the features which caused the separation in urine data between the two groups a number of metabolites were identified and summarised in table 5-10.

Table 5-11 - List of features assigned from urinary HILIC-MS data from S-plot of the OPLS-DA models between term at birth and preterm at term equivalent age			
Potential Metabolite ID	**Mz/rt**	**Group Observed**	**Level of assignment**
3-hydroxy-3-(3-hydroxyphenyl) propanoate	261/39	Term (ESI –ve)	2
Tetrapeptide	459.2/65	Term (ESI –ve)	3
5-(2-Methylpropyl)tetrahydro-2-oxo-3-furancarboxylic acid	207.1/103	Term (ESI –ve)	2
Tyrosol-4-sulfate	239/132	Term (ESI –ve)	2
Prolyl Threonine	215.1/176	Term (ESI –ve)	3
Unknown 1	361.1/274	Term (ESI –ve)	4
Asp Ile Val Tyr	507.2/301	Term (ESI –ve)	3
Acetylhomoserine	160.1/345	Term (ESI –ve)	2
Oxyquinoline	144/42	Preterm(ESI –ve)	2
4-Cresol	107/58	Preterm (ESI –ve)	2
4-Pyridoxic acid	182/63	Preterm (ESI –ve)	2
Lys-Trp-OH	439.2/346	Preterm (ESI –ve)	2
Hydroxyproline	130.2/364	Preterm (ESI –ve)	2
Histidine	154.1/528	Preterm (ESI –ve)	2+
Pantothenic Acid	218.21/218	Preterm (ESI –ve)	2+
Creatinine	136/174	Term (ESI +ve)	2+
Unknown 2	385.1/276	Term (ESI +ve)	4
Unknown 3	201.2/356	Term (ESI +ve)	4
Unknown 4	370.2/385	Term (ESI +ve)	4
Unknown 5	377.3/414	Term (ESI +ve)	4
Proline	116.1/514	Term (ESI +ve)	2
Unknown 6	159.1/569	Term (ESI +ve)	4
Unknown 7	163/98	Preterm (ESI +ve)	4
Pantothenic Acid	221.1/216	Preterm (ESI +ve)	2+
1-*N*-methylnicotiniamide	137/332	Preterm (ESI +ve)	2

For the positive ionisation mode, infants who were born at term separated higher concentration of the metabolites including creatinine and proline. As expected, from the results of the ^{1}H NMR data urinary, creatinine is positively correlated to infants born term. For the positive ionisation mode, infants who are born preterm are related to the metabolites 1-*N*-methylnicotinamide and pantothenic acid.

For the negative ionisation mode, infants who are born term separate due to metabolites including 3-hydroxy-3-(3-hydroxyphenyl) propanoic acid-o-sulphate, 5-(2-Methylpropyl)tetrahydro-2-oxo-3-

furancarboxylic acid, Tyrosol-4-sulfate, acetylhomoserine as well as a number of different polypeptide chains. For the negative ionisation mode, infants which are born preterm separate due to metabolites included 4-cresol, oxyquinoline, histidine, 4-Hydroxy-D-proline, 4-Pyridoxic acid.

5.4.4. IDENTIFYING BILE ACID DIFFERENCES IN INFANTS USING TARGETED MASS SPECTROMETRY PROFILING OF FAECAL SAMPLES

Targeted bile acid profiling using UPLC-MS on faecal samples from infants born term and preterm was performed to further investigate the bile acid composition. Uni-variate statistical analysis was executed on the integrated bile acid integrated peaks to determine any difference between bile acid compositions in infants.

Interestingly - even though previous ^{1}H NMR data showed a separation in term samples from birth due to bile acids - targeted profiling showed no differences in bile acid composition between term versus preterm faecal samples. This may highlight that the global compositional profile of bile acids is different between classes but there are no specific bile acid species differences. Furthermore, meconium samples have a higher concentration of bile acids which may have caused the strong separation in previous models.

Table 5-12. List of bile acids and p-values from univariate statistical analysis on integrated bile acid data from faecal samples taken from term versus preterm infants

Number	Bile Acid Name	p-value
1	3,7,12 Taurodehydrocholic acid	0.45
2	Ursodeoxycholic acid	0.29
3	Ursocholanic Acid	0.988
4	Taurohyodeoxycholic Acid	0.7
5	Taurohyocholic Acid	0.11
6	Taurodeoxycholic Acid	0.25
7	Taurocholic Acid	0.14
8	Taurochenodeoxycholic Acid	0.23
9	Tauro-ursodeoxycholic Acid	0.9
10	Tauro-ursocholanic Acid	0.19
11	Tauro-ß Muricholic Acid	0.91

12	Tauro-a Muricholic Acid	0.65
13	Tauro omega-Muricholic Acid	0.77
14	omega Muricholic Acid	0.85
15	Murocholic Acid	0.34
16	Lithocholic acid	0.26
17	Isolithocholic Acid	0.94
18	Isodeoxycholic Acid	0.25
19	Hyocholic acid	0.49
20	Glycoursodeoxycholic Acid	0.21
21	Glycohyocholic Acid	0.2
22	Glycodeoxycholic acid	0.06
23	Glycochenodeoxycholic Acid	0.28
24	Deoxycholic Acid	0.25
25	Cholic acid	0.21
26	Chenodeoxycholic Acid	0.33
27	a-Muricholic	0.212
28	9(11), (5ß)-Cholenic Acid-3a-ol-12-one	0.55
29	8(14),(5ß)-Cholenic Acid-3a, 12a-diol	0.26
30	5ß-Cholenic Acid-7a-ol-3-one	0.24
31	5ß-Cholanic Acid-3ß, 12a-diol	0.45
32	5ß-Cholanic Acid-3a, 6a-diol-7-one	0.86
33	5a-Cholanic Acid-3, 6-dione	0.12
34	5-Cholenic Acid-3ß-ol	0.9
35	3a-Hydroxy-12 Ketolithocholic Acid	0.29
36	3,7,12 Dehydrocholic acid	0.47
37	3,7-Diketocholanic Acid	0.39
38	3 Dehydrocholic Acid	0.19
39	12 Dehydrocholic Acid	0.62
40	3a-OH-7,12-Diketocholanic Acid	0.66

5.4.5. COMPARATIVE METATAXONOMICS USING 16S rRNA ANALYSIS OF INFANT FAECAL SAMPLES

DNA extraction and sequencing using Illumina technologies was applied to examine the differences in gut microbiome between infants born term or preterm. Extended error bar plots were constructed in STAMP, (Parks et al., 2014) from the MiSeq data and associated confidence intervals with bootstrapping in order to indicate the mean difference for each phyla, class and family in the

infant faecal samples generated. Information on samples used for this analysis can be found in section 5.3.1.

5.4.5.1. Comparison of the gut microbiome composition in infants born term vs. preterm

Gut microbiome composition is indicated as different between infants born term (n=130) and preterm (n=164), in particular bacteria from the class of Gammaproteobacteria, which is from the Proteobacteria phylum, are highly significant in preterm born samples compared to term born infants. More specifically, at a genus level, preterm infants have significant proportion of an unclassified *Enterobacteriaceae, Finegoldia* and an unclassified *Staphylococcus* (Figure 5-19a).

Whereas, bacteria from the Actinobacteria and Bacteroidetes phyla are significantly higher in samples from infants born at term. Furthermore, term born infants have higher amounts of bacteria from the Alphaproteobacteria which is also from the Proteobacteria phyla similar to the preterm infants. At a genus level, term infants have more significant proportion of *Streptococcus*, *Escherichia/Shigella*, *Lactobacillus*, *Clostridium*_XI, *Collinsella* and *Bacteroides* (Figure 5-19b).

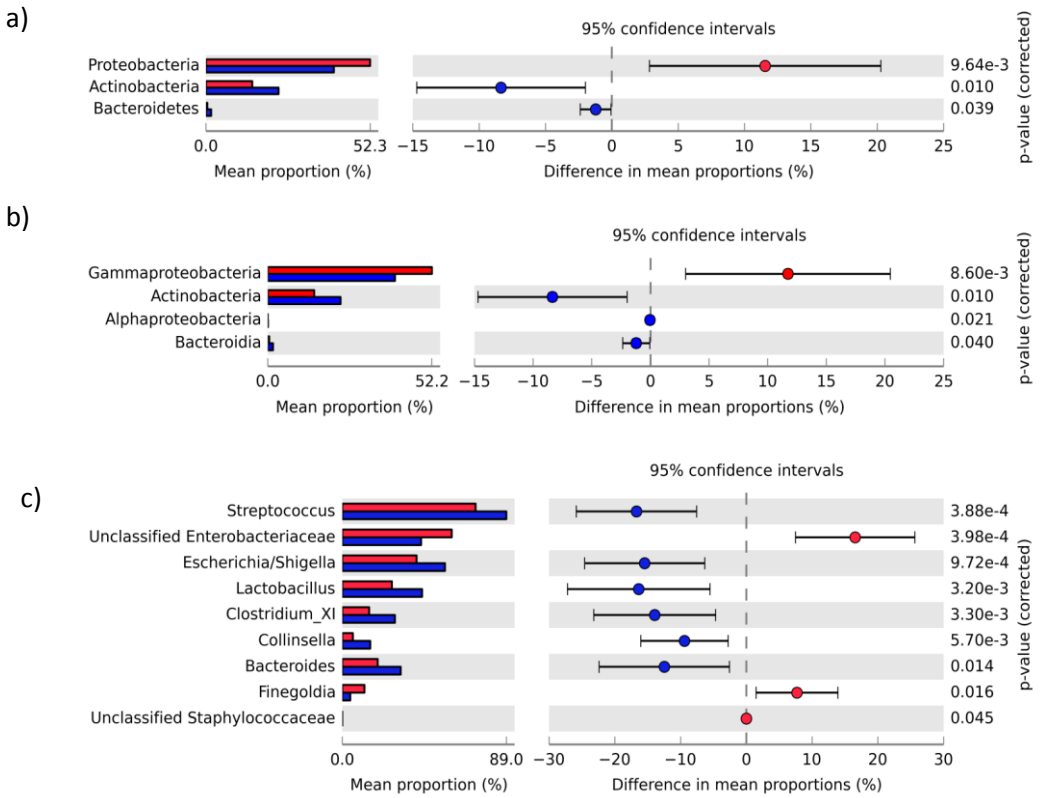


Figure 5-19: Extended box plots constructed from MiSeq DNA sequencing data from faecal samples of infants born term (blue) versus preterm (red) a) phylum, b) class, c) genus

5.4.5.2. Comparison of the gut microbiome composition in infants born term at birth vs. preterm at term equivalent age

Gut microbiome composition is indicated as different between samples from term infants at birth (n=28) compared to preterm infants when they reach term equivalent age (postnatal age > 37 weeks; n=83). In particular bacteria from the Bacteriodetes phylum (Figure 5-20a), including the Bacteriodia and Clostridia classes, are highly significant in term infants samples at birth compared to preterm born infants at term equivalent age (Figure 20b). More specifically, at a genus level, term infants have significant proportion of *Clostridium_XI, Bacteriodes, Methyloacterium*, unclassified *Comamonadaceae*, *Allobaculum*, *Escherichia*/*Shigella* and *Bradyrhizobium* (Figure 5-20c). Whereas,

samples from preterm infants at term equivalent age, have higher proportions of unclassified *Enterobacteriaceae*.

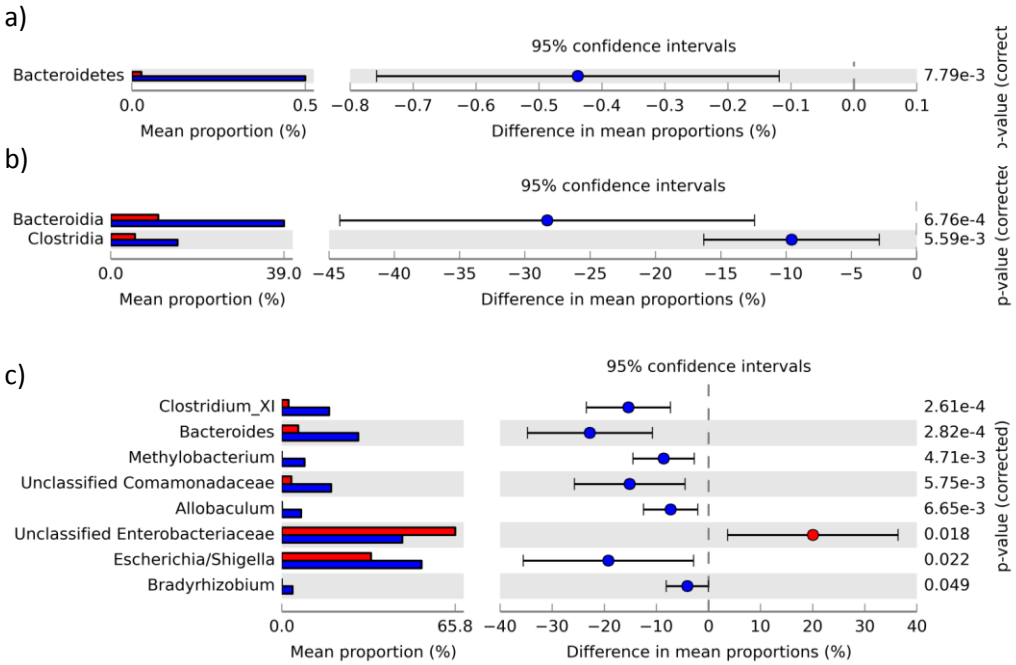


Figure 5-20: Extended box plots constructed from MiSeq DNA sequencing data from birth faecal samples of infants at term (blue) versus preterm at term equivalent age (red) a) phylum, b) class, c) genus

5.4.5.3. Comparison of the gut microbiome community in infants born term vs. preterm at 3 months postpartum

After three months postpartum, the gut microbiome composition is indicated as different between samples from term (n=31) compared to preterm infants (n=31). In particular bacteria from the Actinobacteria phylum are highly significant in term infant samples at 3 months compared to preterm born infants (Figure 5-21a). More specifically, at a genus level, term infants have significant proportion of Actinomyces which is part of the Actinobacteria phylum. Whereas *Clostridium*_XVIII and *Streptococcus* compared to preterm infants (Figure 5-21c).

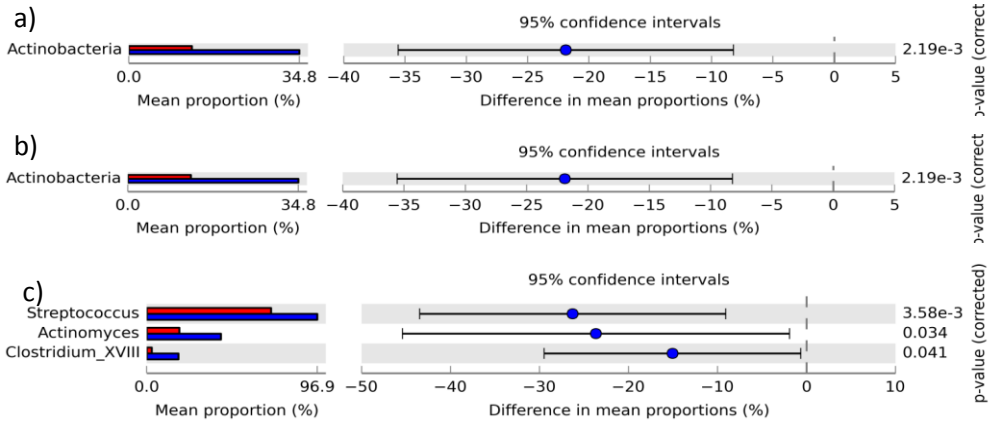


Figure 5-21: Extended box plots constructed from MiSeq 16S rRNA sequencing data from faecal samples of infants born term (blue) versus preterm (red) taken at three months postpartum a) phylum, b) class, c) genus

5.4.5.4. Comparison of the gut microbiome community from term born infants born through different modes of delivery

The gut microbiome composition is indicated as different between samples from term born infants born through caesearean section (n=71) compared to infants born vaginally (n=95). In particular bacteria at a genus level, vaginally delivered infants have significantly higher proportion of Escherichia or Shigella in their stool compared to infants which are born through caesarean section (Figure 5-22).

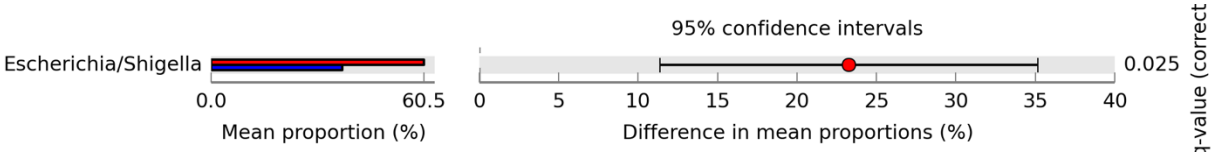


Figure 5-22: Extended box plots of genus constructed from MiSeq 16S rRNA sequencing data from faecal samples of infants born through vaginally delivery (red) versus caesarean section (blue).

5.5. DISCUSSION

The current study has shown that there are clear differences in both metabolic profiling data using different platforms (e.g. ^{1}H NMR and UPLC-MS) between infants born at term age (>37 week of gestation) compared to infants born preterm. Infant biofluids are a valuable resource to understand the development of the new-born infant after birth as well as during the first few months of life.

This chapter demonstrates the different developmental trajectory between infant born term and preterm using metabonomics, metagenomics and microbiomics. Key metabolites involved in developing renal function, nutritional influence as well as gut microbiome interactions can clearly be recognised.

5.5.1. METABOLIC AND MICROBIOMIC PROFILING TO CHARACTERISE BIOFLUIDS OF INFANTS BORN TERM VS. PRETERM

From these results, it is clear that by using both NMR and HILIC-MS it can show differences in urine and faecal samples from infants born term and preterm at all time points. In particular, urinary metabolic profiling using NMR for term born infants correlated highly to choline specific metabolites (e.g. betaine, dimethylamine and dimethylglycine). Infants' development continues after birth which is reiterated in these results as dietary choline isn't being fully utilised due to their immature gastrointestinal and kidney system (Foxall et al., 1995; Trump et al., 2006; Nicholson et al., 2012). Betaine could be released due to the immaturity of the infant kidney, which is still developing postpartum even in term born infants. Infants born term are also highly associated to TMA in their urine; TMA is produced by the microbial bioconversion of choline and may be related to the developing gut microbiome (Chiu et al., 2016). Furthermore, the presence of TMA also indicates the gut microbiome successfully utilising choline from the diet. Furthermore, term infants' urine is correlated to homogentisic acid which is an intermediate of the metabolic breakdown of tyrosine

and phenylalanine; these metabolites are related to the developing gut microbiome (Marcobal et al., 2012).

From the urinary HILIC MS data it is in agreement with the urinary NMR data; also showing a higher concentration of di- and poly- peptides, which suggests a lack of full protein digestion by the digestive system. Term infant urine is also associated with linoleoyl-glycerophosphocholine, a common phospholipid bound choline found in human breast milk (Holmes-McNary et al., 1996, Ilcol et al., 2005). Preterm infant's urine also has high concentrations of sugar in their urine (e.g. mannitol, fructose as well as unidentifiable sugars and carbohydrates) which may be due to a lack of absorption of the simple molecules in their immature organs or due to higher proportion of sugars in preterm nutritional feed in early life (Radmacher et al., 2016). Preterm infants also have the metabolite *N*-acetylgalactosamine-6–sulphate in their urine, this metabolite has previously been recognised in urine (Hopwood and Elliot, 1985). This sulphated *N*-acetylhexosamines has been identified as an important part in sulphated glycosaminoglycans, glycoproteins and sugar nucleotides. *N*-acetyl glycoproteins are commonly found in breast milk so this could have been sulphated by the infants before excreted in the urine (Tannock et al., 2013). Histidine is an amino acid which cannot be synthesised by the body and is taken up through our nutrition; it is known that a deficiency of histidine in the first 6 months of life can cause a decreased weight gain as well as the potential to cause dermatitis (Synderman et al., 1963). Hao et al., (2015), noted that there was a lower concentration of histidine, as well as other amino acids, found in the urine of preterm infants. They concluded that the lack of metabolic enzymes (Vockley et al., 2006) or lack the ability to absorb amino acids by the body was the cause.

Faecal water ^{1}H NMR metabolic profiles from infants born at term are highly associated to bile acids species. The distinction between bile acid species is very small, depending on the presence and absence of hydroxyl groups on positions 3, 7, and 12 and therefore many bile acid species are found at similar, if not the same, chemical shifts on an NMR spectra. This said however, cholate, a major

primary bile acid, was identified from the ^{1}H NMR faecal water spectra and found to be highly correlated to term born infants; this bile acid is produced in the liver and usually conjugates with glycine or taurine (Begley et al., 2005). Bile acids have a role in digestion as it emulsifies and solubilises fats for intestinal absorption (Houten et al., 2006). Bile acids also repress the growth of microbiome by solubilising the membrane lipids, causing lysis and cell leakage. Faeces produced soon after birth is called the meconium; it is composed of materials which are ingested during the time in the womb. It typically consists of epithelial cells, hair, mucus and bile as well as water and therefore has a different consistency to stool later in life. With the high proportion of meconium samples in term infants compared to preterm infants, this separation could be caused by term birth samples alone.

From the ^{1}H NMR urine data, *N*-1-methylnicotinamide (NMN) is related to preterm birth; this is an organic cation synthesized from niacin and tryptophan and is one of the end products in the nicotinamide pathway. It typically is excreted in the urine and can be used as a marker of proximal tubular necrosis and renal injury (Trump et al., 2006). Preterm infants renal function further develops outside of the womb, so this metabolite is likely to be part of this process.

Urinary HILIC-MS data shows that preterm infants are related to the metabolite hydroxynicotinic acid; this metabolite is broken down by bacteria, specifically *P.aeroginosa*, by metabolising nicotinic acid. This metabolite can be used to diagnose urinary tract infections from urine samples and as preterm infants have many different complications from birth including infections, this diagnosis is not surprising (Gupta et al., 2005; Bauer et al., 2003).

Additionally, urinary HILIC-MS data shows preterm infants are highly correlated with 7-methylguanine. This metabolite, 7-methylguanine, is a methylated version of the nucleoside guanosine and may reflect an increased turnover of nucleic acids related to tissue modelling and development in early life. Infants excrete between six and ten times more RNA catabolites compared

to adults and this correlates to the higher protein turnover rates in new-born and preterm infants compared to adults (Sander et al., 1986).

Furthermore, infants born preterm separate due to metabolites included histidine and ascorbic acid. Ascorbic acid is one form of vitamin C; this vitamin isn't stored by the body so humans require it as part of their daily nutrition. Preterm infants usually are given a high amounts of calcium, phosphorus as well as ascorbic acid to protect them from bone deformities due to prematurity (Wojcik et al., 2009). This may be the cause for an excessive amount of ascorbic acid excreted in the preterm urine (Schell-Feith et al., 2000, Darlow et al., 2005). Furthermore, preterm infants' faecal water metabolic profile is highly associated with thiamine, also known as vitamin B1, which plays a key role in the metabolism of glucose and converting carbohydrates and fats into energy that the body utilises.

Faecal water ^{1}H NMR profiles of preterm infants are highly associated with glycerol compared to term infants. This is usually an exogenous metabolite and used widely in pharmaceutical formulations which preterm infants could potentially be subject to, for a range of different disorders. However, glycerol has previously been recorded in adult human faecal samples as an endogenous metabolite when comparing it across species (Saric et al., 2007) and is thought to be important in lipid metabolism (Fanos et al., 2013).

In terms of the developing gut microbiome, differences between the two groups were observed. Bacterium from the *Actinobacteria* and *Bacteroidetes* phyla were significantly higher in samples from infants born at term. *Bacteroidetes* species of bacteria are known to break down dietary choline into TMA, which supports the metabolic data from this group (Romano et al., 2015). From previous research, it has been noted that lower levels of *Bifidiobacterium* and *Bacteroides* in infants born preterm compared to term born infants (Stark et al., 1982).

Furthermore, term infants are highly associated to *Lactobaccillus* species, this bacterium is known to reside in breast milk and transferred to the infant. Term infants are feeding on more breast milk

than preterm infants which can be recognised in the high abundance of this bacterium as well as the metabolites found in their urine and faeces; including choline and linoleoyl-glycerophosphocholine.

Infants born preterm usually have a reduced diversity and higher levels of pathogenic bacteria; this is mainly due to their time spent in hospitals so that they are exposed to these pathogens at a higher level as well as they are subject to potential antibiotic treatments causing dysbiosis (Cong et al., 2016). Preterm infants in this study are more associated with *Proteobacteria*, it has been noted previously that these bacteria are related to the use of antibiotic treatments in early life (Fouhy et al., 2012; Cong et al., 2016).

Preterm birth is highly associated with necrotising enterocolitis (NEC); an inflammatory intestinal disorder thought to be influenced by the gut microbiota. Unfortunately, for this project there was no additional information on whether these infants contracted NEC during this study or the mode of delivery for these infants. It is known that an increased abundance of *Gammaproteobacteria* is seen in NEC patients, suggesting that this class of bacteria may be predictive of this disease (Wang et al., 2009). Preterm infants, also, have a higher association to *Staphylococci* and *Finegoldia* bacteria, which is usually found on the skin (Groer et al., 2014). This could be due to these infants being born through caesarean section and being exposed to the skin of the mother, especially during breast feeding. These bacterium are commonly found on the skin and regarded as commensal; however, they are also opportunistic pathogens and cause infections in open wounds.

5.5.2. METABOLIC AND MICROBIOMIC PROFILING TO IDENTIFY DIFFERENCES IN INFANTS AT TIME OF BIRTH

As we have discussed before, term and preterm infants can be anatomically and physiologically different to each other due to the difference in gestational age. Understanding the difference in metabolic profiles between these infants helps to understand the needs of a premature baby metabolically.

Term infants have higher concentrations of pseudouridine in their urine; the isomer of the nucleotide uridine found in RNA is a degradation product of transfer ribonucleic acid (tRNA). This metabolite has also been attributed as a metabolic marker for kidney nephrotoxicity in new born rats so may be an indication into the developing kidneys from birth (Hanna et al., 2013; Fanos et al., 2013).

Furthermore, term infant's urine correlates to trimethylamine (TMA) which is related to choline metabolism (Trump et al., 2006). When there are high carnitine levels; bacteria from the gut microbiome degrade carnitine in the body to make TMA. This is an interesting find, as until recently infants were thought to be sterile inside the womb, but if it would be possible to attribute this metabolite to bacterial co-metabolism this would mean that bacterial activity was occurring in the womb. That said however, in my previous investigations (chapter 3) it has been noted that metabolic profiles of urine at birth is very similar to that of the mothers. Therefore, this metabolite could have been passed from mother to foetus in the womb through the placenta. With this in mind, this metabolite is related to term infants and not in preterm, suggesting that there is a lack of choline exchange from mother to infant in preterm infants or this exchange occurs later in gestation.

From the urinary HILIC MS data, infants which are born term separate from preterm due to metabolites including dipeptides and amino acids, as well as citrate and homocarnosine. Homocarnosine is related to histidine and the brain-specific dipeptide, gamma-aminobutyric acid (GABA) as well as acetylhomoserine (Pearl et al., 2006). The presence of polypeptide chains in the urine could be to do with the lack of digesting diets fully. Furthermore, similar to term samples, urine samples were taken at birth and could be related to mother's metabolic profile. These are the same features which caused the separation in the previous model which is to be expected as the only samples run by HILIC-MS from term born infants were obtained at birth only.

Furthermore, term infants are related to citrate, this metabolite is an intermediate in the TCA cycle. This cycle is very important for humans as it produces energy through the oxidation of acetyl-CoA in

the mitochondria (Bender and Mayes, 2012). This metabolite has previously been seen in term infant urine when compared to preterm (Foxall et al., 1995).

Faecal water NMR metabolic profiles from term infants at birth are highly associated with bile acids as well as an unknown metabolite which is highly associated to this region (δ0.65-0.85) suggesting that it could be a bile acid conjugated to taurine (Duarte et al., 2008). As discussed previously, meconium samples are composed of very different molecules compared to stool obtained from later postpartum samples. When an infant is born, the gastrointestinal tract sheds its protective layer consisting of bile, epithelial cells and hair which makeup the meconium. It is evident from this project that meconium samples have higher concentrations of bile acids and therefore a different faecal metabolic profile compared to samples taken at a later date. From targeted analysis of bile acids in these samples, there were no significant differences seen in infants born term compared to preterm which would suggest that there is an overall quantitative difference in bile acids rather than specific species. Furthermore, the gut microbiome, in particular *Bacteroides* species, can utilise the taurine conjugated bile acids by enzymatically deconjugating it. Thus releasing the amino acid side chain to use it as a source of carbon, nitrogen and energy sources (Begley et al., 2006).

Urinary NMR metabolic profiles from preterm infants at birth are highly correlated to threonine, 1-*N*-methylnicotinamide and myo-inositiol as well as an unknown metabolite which was tentatively assigned as α-aminoisobutyrate. Threonine is a glucogenic amino acid, which is found in proteins, and is needed to generate glycine and serine for the production of muscle tissue; it has been previously associated with infants born small for gestation age (Moltu et al., 2014).

Myo-insitol plays an important role in the structural component of lipids and various phosphates as well as a myriad of different roles within the body (Foxall et al., 1995). Inositol is produced by the human body from glucose and used for many cellular functions; it is also thought to be a marker of glucose intolerance (Mahan 2011). Preterm infants have altered glucose metabolism and potentially have developed a hypo-insulinemic condition (Dessi et al., 2014). Elevated abundance in preterm

infants of myo-inositol may also suggest differences in the synthesis of membrane phospholipids which act as lung surfactants (Diaz et al., 2015). Furthermore, myo-inositol is also found in breast milk and usually fully absorbed into the blood.

However, it is seen to be increased in diabetic or renal failure patients due to the lack of absorption in the kidneys. Preterm infants have immature renal systems so probably lack the efficiency of taking up all the myo-insitol before it can be excreted. Furthermore, α-aminoisobutyrate and 1-*N*-methylnicotinamide are related to the renal function development after birth as well (Trump et al., 2006).

Urinary HILIC-MS data from infants born preterm separate due to metabolites including ascorbic acid and glutamine as well as hydroxynicotinic acid. Glutamine is an amino acid found at high concentrations in human breast milk. It has been previously noted that glutamine in cord blood positively correlated with birth weight in infants, which indicates the difference in anabolic rate that low birth weight infants have compared to term born infants (Ivorra et al., 2012). Nicotinic acid metabolites have been linked to urine from children with ASD; preterm infants are at a higher risk of developing conditions including ASD (Yap et al., 2010).

Preterm faecal metabolic profiles taken from birth was highly correlated to carbohydrates and sugars seen at NMR chemical shifts between 3-5ppm, which have many overlapped resonances from glucose, other sugars, polyols, *N*-methylated compounds and amino acids. Preterm infants are not fully developed, especially their digestive tract, therefore simple sugars from their diet are not fully absorbed and are lost in their faeces. Further to this, preterm infants stool samples are more associated to the amino acids; alanine and phenylalanine. Phenylalanine is naturally found in breast milk and included in nutrient supplements for preterm infants (Atzori et al., 2011). Therefore, again, due to the immaturity of the digestive tract, this amino acid is not fully absorbed into the body or there is an excess introduced into the body from the diet.

5.5.3. METABOLIC AND MICROBIOMIC PROFILING TO IDENTIFY DIFFERENCES BETWEEN TERM INFANTS AT BIRTH VS PRETERM AT TERM EQUIVALENT AGE

Infants which are born preterm are deemed term equivalent age when they have reached 37 weeks' gestation in the weeks spent postpartum. By comparing term infants birth samples and samples taken from preterm infants when they reach term equivalent age (37-41 gestation week) it gives a clearer idea as to how the preterm infants have developed compared to infants which are born at term.

From the urinary NMR data, there was no significant difference in urinary metabolic profiles between term samples taken at birth compared to preterm samples taken at term equivalent age (Supplementary Figure S5). However, there was a significant difference between urinary global profiles using HILIC-MS. Infants born term separate due to metabolites including creatinine. As expected from previous models of NMR data, creatinine is correlated to infants born term. This metabolite is related to muscle mass in humans and even though infants born preterm have reached term equivalent age in this model and we would expect them to have a higher muscle mass, this would suggest that term born infants still have a higher muscle mass from birth.

Furthermore, infants born term separate due to nutritional metabolites including 3-hydroxy-3-(3-hydroxyphenyl) propanoate, 5-(2-Methylpropyl)tetrahydro-2-oxo-3-furancarboxylic acid, tyrosol-4-sulfate, as well as a number of different polypeptide chains and proline. 3-hydroxy-3-(3-hydroxyphenyl) propanoate is a metabolite from the breakdown of phenylalanine which comes from nutritional sources or as these samples were taken at birth these metabolites could be related to the maternal metabolic profile which has been exchanged through the placenta in the womb (Kumps et al., 2002). 5-(2-Methylpropyl)tetrahydro-2-oxo-3-furancarboxylic acid is typically found in milk products and tyrosol-4-sulphate is a common polyphenol metabolite found in dietary sources including infant formulas (Pande and Akoh, 2015; Neveu et al., 2010). Additionally, if these samples

were taken after the first feed from the mother, this would alter the metabolic profile, which supports the need to standardise and be aware of time records of samples in infants.

Infants born preterm separate due to metabolites including 1-*N*-methylnicotinamide and pantothenic acid. 1-*N*-methylnicotinamide was also a metabolite which caused the separation of classes in the NMR data. As previously mentioned, this metabolite is an organic cation synthesized from niacin and tryptophan and is one of the end products in the nicotinamide pathway which can be excreted in the urine (Trump et al., 2006). Pantothenic acid and 4-pyridoxic acid, which were also identified by HILIC-MS, are also known as vitamin B_5 and a catabolic product of vitamin B_6 respectively. These metabolites are not synthesised by the human body and is found in breast milk as well as synthesised from many different species bacteria which reside in the gut (Hooper et al., 2001, Kau et al., 2011). Breastmilk concentration of this metabolite increases over time from mothers of full time as well as preterm mothers (Ford et al., 1983). Therefore, it is reasonable that preterm infants which have been developing outside the womb, are exposed to more bacterial co-metabolism as well as ingesting breastmilk or exposed to other dietary factors, will have a higher concentration of this metabolite in their system.

Preterm infants separate due to 4-cresol; this metabolite is a phenol derivative and a microbial co-metabolite which is produced by the fermentation of protein by the gut microbiome (Maitre et al., 2014). This gives further evidence for the microbial activity in the preterm infant's intestinal system is contributing to their metabolic health.

Infants which are born preterm separate due to the metabolite histidine, this amino acid used in the synthesis of proteins and known to be essential for infants up until 6 months. (Synderman et al., 1963) This metabolite has also previously been seen to be higher in preterm infants when compared to term infants (Hao et al., 2015). Preterm NMR metabolic profiles are highly associated to amino acids (including alanine, leucine, lysine and tyramine) compared to term infants. Even at a later stage of development preterm infants are still not fully developed, therefore amino acids from their diet

are not fully absorbed and are lost in their faeces. Tyramine is a monoamine compound which is derived from tyrosine and related to phenylalanine; it is naturally found in breast milk, as well as contained in many parenteral nutritional supplements.

Furthermore, infants born preterm correlated to hydroxyproline, this metabolite is a major component of collagen, the main protein to produce the extracellular space in tissues. This metabolite is regularly excreted in urine and is thought to be a reliable indicator of collagen turnover and the concentration excreted is related to rapid growth (Younoszai, et al., 1969). As this metabolite is higher in preterm infants it would suggest that these infants are rapidly growing compared to their term urine taken at birth.

Faecal NMR metabolic profiles also showed a significant separation according to class; there is a higher correlation of bile acid in the term infant samples compared to the preterm babies, which was seen in the previous models in this chapter. The bile acids in term babies are associated with the different composition of the meconium. Bile acids also repress the growth of microbiome by solubilising the membrane lipids, causing lysis and cell leakage, this may be an important factor to protect the foetus from bacterial growth in the developing intestine whilst in the womb.

The extended box plot shows that the phylum level, *Bacteroidetes* is highly associated with term infants at birth compared to preterm infants at term equivalent age. *Bacteroidetes* and *Firmicutes* are the most prominent phyla in the adult gut microbiome, *Bacteroidetes* in infant samples are related to breast milk as previously mentioned. Delving deeper into the differences of genus between term and preterm, there are many different species of bacteria associated to the term infant. This supports the notion that term infants have a much higher diversity of bacterium in their gut compared to preterm infants. Additionally, *Allobaculum*, which is part of the *Firmicutes* phyla, is associated to term infants and found in healthy infants and not influenced by the mode of delivery (Rutayisire et al., 2016). Interestingly, bacterium genera, *Methylobacterium* and *Bradyrhizobium* are related to term infants. This bacterium is usually found in soil and tap water, even though they have

been found in human microbiome samples, including the ocular surface, they have not been recognised in infant studies before (Dong et al., 2011).

Early life exposures can affect the development of the gut microbiome, especially preterm birth and antibiotic use can have long term effects on the microbial diversity. Infants born preterm have a higher association to the aerobic, *Enterobacteriaceae* at the time of term equivalent age. It has been discovered that this particular bacterium is related to the hospitalisation in a NICU for infants (Parm et al., 2011). Nutritional factors are believed not to be related to the colonisation of this bacterium but it is potentially related to antibiotic use.

This data shows that even after preterm infants have reached term equivalent age, where they have been exposed to the environment longer compared to the term infants where samples were taken in the first day still have a perturbed microbiome development.

From previous investigations within this thesis, it has been repeatedly noted that birth samples from infants have a different composition compared to samples taken later postpartum. Therefore, modelling term birth samples against later samples from preterm infants was evident that there were going to be disparities. Consequently, it was decided to produce another model where samples from term infants were from weeks 1-2 postpartum rather than birth samples. Another benefit of this model is that metabolites contributed from maternal metabolism aren't regarded in the analysis and the focus' is purely on the infant metabolic profile. Preterm samples have relatively elevated concentrations of *N*-acetyl glycoproteins, which are found in breast and formula milk so this metabolite is likely to have come from the diet.

Interestingly, the previous model when term birth samples were modelled, preterm samples were associated with tyramine; however, in this updated model, when term samples were taken at one week postpartum, tyramine is associated with term infants conversely. Tyramine is the break down product of tyrosine, a common amino acid which can be executed by lactic bacteria (Marcobal et al.,

2012). This result suggests that after a week postpartum, lactic acid bacteria colonising term infants are utilising more tyrosine for nutrients at this time compared to at birth and compared to preterm infants, when they would have had a much lower gut microbiome diversity.

5.5.4. METABOLIC AND MICROBIOMIC PROFILING TO IDENTIFY DISPARITIES BETWEEN INFANTS AFTER THREE MONTHS POST-PARTUM

After three months of development outside of the womb and interactions within the environment, infants born preterm have been exposed firstly to the hospital environment, which is staggeringly different to their term born counterparts. Equally, infants born preterm are usually subject to many different therapies to treat with their numerous medical conditions. These first experiences can have a huge effect on the developing body of the new born infant and thus has the potential to dramatically change their metabolic profile; therefore, by comparing metabolic profiles at three months after birth these differences have been assessed.

Urinary NMR data from term infants at three months postpartum is highly associated with betaine, dimethylglycine, dimethylamine. As discussed previously, these metabolites related to the metabolism of dietary choline where there may be an imbalance of the utilisation of choline from the immature gut. Furthermore, betaine could be released due to the immaturity of the infant kidney (Foxall et al., 1995). *N*-acetylglutamate is also highly associated with this class, this metabolite is related to breast milk that the infants diet consists of (Andreas et al., 2015).

Preterm infant's urine at three months postpartum is highly associated with mannitol, fructose and a species of bile acid. Mannitol is a monosaccharide and is used widely in medical use; it is used for the reduction of intracranial pressure as well as prevents swelling of the endothelial cells in the kidney. Furthermore, it can be used for the evaluation of intestinal permeability of the small intestine; many preterm infants have immature gut maturation and as a result have underdeveloped gut mucosal barrier. Fructose, which is also a monosaccharide, is shown to be the principal sugar of

fetal blood while it is found in minute amounts in the maternal blood (Trindade et al., 2011). The elevated concentration of this sugar in the preterm infant's urine could be that the preterm infant is continuing to produce this sugar as an alternative pathway in glucose pathway which they would have done as a fetus.

The preterm infant urine has higher concentrations of 1-*N*-methylnicotinamide (NMN), an organic cation which was identified in previous models. As it is still causing a separation between preterm and term samples it could mean that the preterm renal function is still underdeveloped compared to term infants. Furthermore, preterm infants had elevated bile acid concentration at this time point compared to term infants. This has previously been recognised and thought to be an over production or clearance of bile acids by the preterm body (Yamato et al., 2001). Although, another reason for this could be that preterm infants are not as efficient at reabsorbing bile acids from the colon and therefore are excreted more readily compared to the term infant.

Lysine is an essential amino acid for protein synthesis and found in protein rich foods, it is also the first limiting amino acid in milk fed mammalian newborns. If the amount of protein in the diet is not adequate, the rate of protein synthesis and growth will be determined by the amount of lysine (van der Schoor et al., 2004). From this model, different feeding types contributed to the metabolic profiles, which is reiterated when investigating the specific metabolites causing this difference. With an elevated relative abundance found in preterm infants, this would suggest that they are being fed a high protein diet or are not utilising all the lysine from their diet and therefore may be subject to slow growth rates.

Conversely, faecal water metabolic profiles at three months' post-partum age was not distinguishable between infants born term compared to preterm born infants. This age dependant similarity in metabolic profiles may be due to the same environmental exposures and diet is the most influential factor on the faecal profile rather than human development.

The gut microbiome is constantly changing in the first few months of life and doesn't completely stabilise until we are toddlers. This is due to the consistent introduction to new environmental exposures (e.g. changes in diet from milk to solid food) as well as the general maturity of the human body. It was observed that bacteria from the Actinobacteria phylum are significantly higher in samples from infants born at term. Other studies have concluded that the main bacterial species found in the infant microbiome are *Bacteroidetes* and *Bifidiobacterium*, the latter is part of the Actinobacteria phylum and supports this result.

At the genus level, term infants are highly correlated to the *Actinomyces* from the *Actinobacteria* phylum as well as *Streptococcus* and *Clostridium* XVIII from the Firmicutes phylum. Actinomyces genus are related to higher fat content found in breast milk, specifically the oleic and cis-vaccenic acid content (Taft et al., 2016). From the metabolic data, urine profiles separate according to the particular nutrition, further linking metabolome and microbiome information as complementary.

Streptococcus bacterium are commonly found in breast milk and as more term infants are on a diet of breast milk compared to preterm infants who are on a combination of parenteral nutrition and milk this result is predictable (Musilova et al., 2014). However, *Clostridium* species are commonly associated with formula fed infants (Penders et al., 2006), which suggests that both *Clostridium* and *Streptococcus* are highly associated with term infants due to the fact that preterm infants are subject to parenteral nutrition and perhaps have a lower diversity of bacterial species in their gut microbiome.

5.6. CONCLUSION

Using multi-analytical platforms provides an extensive range of biochemical and microbiome related information. The findings from this study show that the metabolic profile of urine and faecal water is strongly affected by gestational age at birth and increasing postnatal age. Specifically, choline metabolism is positively correlated to infants born at term whereas other dietary metabolites can be

found in preterm infants, including amino acids and other organic acids. This could be a key finding for future dietary management in the developing preterm infant. Interestingly, no disparities in SCFAs were recognised in this study between term and preterm infants; these compounds are produced by the fermentation of polysaccharides by the gut microbiome. This is intriguing as the gut microbiome has been documented as different between the two groups but this suggests that their functionality of utilising these as a nutrient source are equal.

Furthermore, there is a large range of metabolites being excreted in term and preterm infant urine which were related to kidney development and potential injury or damage. This would suggest that the development of renal function is different in term and preterm infants from birth. Further investigation into the effects of these different metabolites have on the body needs to be a priority.

CHAPTER 6

GENERAL DISCUSSION

The investigations that form this thesis have determined the metabolic development trajectory of infants in their first two years of life. Using ^{1}H NMR, UPLC-MS and metagenomics profiling techniques to generate urine and faecal samples from healthy neonates and infants shows the breadth of knowledge we can gain from a further understanding these profiles. Characterising the metabolic profile is critical in the role to understand the normal development trajectory in the first years of life and can be used as a benchmark for future research studies.

6.1 KEY FINDINGS

From the results in this thesis the key findings revolve around establishing clear differences in urinary and faecal metabolic profiles from developing infants, influenced by environmental exposures early in life; including gestational age, mode of delivery and nutrition.

Firstly, my aim was standardising protocols and standard operating procedures for infant biofluid collection and metabolic profiling using mainly ^{1}H NMR analysis, followed by acquisition of urine and faecal profiles over the first two years of life. One of the most interesting findings was the striking similarity of the urinary metabolic profile from infants at birth compared to the maternal urine,

which has not been reported previously (Jackson et al 2016). The age variability in the urinary metabolic profile was key in the focus of this PhD as it was evident that the metabolic trajectory of the infant metabolome, especially the faecal metabolic profile, has not been properly examined particularly with respect to characterizing the metabolic evolution over regular increments.

Two infant studies were used to further our knowledge of metabolic profiling from this critical developmental stage of life. In chapter 4, faecal samples taken in early life (≤120 days postpartum) discriminate according to early life exposures: including gestational age, birth weight, mode of delivery, nutrition, as well as BMI status at 1 and 2 years of age. Interestingly, faecal samples taken at 1 and 2 years postpartum showed very little discrimination between classes (gestational age, birth weight, mode of delivery and nutrition) except for BMI status. This suggests that the differential effect of birth weight, gestational age, mode of delivery and nutrition on the faecal metabolic profile doesn't persist once the infant is over 1 years old.

For chapter 5, a deeper exploration of the early life metabolic differences between infants born term and preterm from birth and until 3 months postpartum. To accompany the faecal metabolomes, which gave insight into the functionality of the gut bacteria, profiling of gut microbiome of infants provided information pertaining to the composition of the gut microbiome. Further evidence for little discrimination for faecal metabolic profile between term and preterm infants after the neonatal period was established in this study, as faecal samples at 3 months also show very little difference between profiles.

Sources of metabolite variation in infant urine and faecal samples from both studies can be assigned to classes of compounds or specific biochemical pathways from these two studies (Table 6-1 and 6-2).

6.2 STANDARDISING INFANT BIOFLUID PROTOCOLS FOR METABOLIC PROFILING

It became apparent at the beginning of this project that infant metabolic profiling studies were not as common as adult studies. Therefore, specific protocols for infant biofluid collection and preparation were not as available (Goodpaster et al 2011). Thus a significant aspect of this project involved the optimisation of protocols before confident evaluations could be accomplished. Being aware of contamination from collection materials (cotton wool and nappies) is essential to the success of an infant metabolic profiling project. For example, propylene glycol was found to be a contaminant coming from collection materials but it can also be produced in the body and by the gut microbiome from glutamine (Jackson et al 2016). Additionally, infant studies come with the difficulty of low volume samples making multiplatform analysis difficult. Evaluating and analysing dilution series in biofluids using ^{1}H NMR, this problem has been addressed and the technique successfully utilised in this project.

Profiling faecal material is a relatively novel tactic for delving deeper into the metabolic profile (Jacobs et al 2008; Saric et al 2007; Chow et al 2015; Martin et al 2014). This biofluid is ideal for further understand the mutualistic relationship between the gut microbiome and host through gaining further knowledge on the functional qualities of this community. Furthermore, similar to urine, this biofluid is ideal for testing as it is a waste product and therefore non-invasive; which is particularly beneficial for infant studies. As mentioned previously, volume of samples from infants (especially preterm born infants) was consistently low and represented a particular limitation in this project, therefore producing a faecal water preparation protocol to get as much information from the single sample using multi-platform techniques was essential for this project. This work was successful especially as infant samples are regarded as precious commodities and often hard to obtain due to sensitivities around sample collection, limited volume and difficulties in obtaining ethics. I would expect that this work will be utilised further in adult studies as it significantly reduced

preparation time as well as preparatory inaccuracy for faecal samples between different analytical methods (NMR and UPLC-MS).

Table 6-1. Summary of the main faecal water discriminatory metabolites and their biochemical pathway from infants at 4 days to 2 years postpartum from the NoMic Study. Key: Amino Acids 3 letter key used.GA- gestation age, MoD – Mode of delivery, BMI – Body mass index, 3'-FSL – 3'fucosyllactose, 4-HPA – 4-hydroxyphenylacetate, TMA – trimethylamine, DMA- dimethylamine, TMAO – trimethylamine-*N*-oxide, N^3-TML – *NNN*-trimethyllysine, 3-HB- 3-hydroxybutyrate, GABA – γ-aminobutyrate. (*overlapped resonances from glucose, other sugars and polyols, *N*-methylated compounds and amino acids)

Class of compound/ biochemical pathway	**Discriminatory metabolites correlated to groups from multivariate statistics (OPLS, OPLS-DA) and time series analysis**						
	Age variability	**Term *vs*.Preterm**	**Birth weight**	**Birth GA**	**MoD**	**Nutrition**	**BMI**
TCA cycle intermediates (e.g. succinate, citrate)	↑ **≥ 120d pp** - succinate	**Term** - ↑ succinate	↑ **HBW**-succinate	↑ **HGA** – succinate			
Sugars and oligosaccharides (e.g. glucose, fructose, mannitol)	↑ **≤ 30d pp** – Sugars*	↑ **Preterm**, Sugars*, 3'-FSL ↑ **Term** Sugars*, glucose	↑ **LBW** -3'FSL ↑ **HBW** – Sugars*, Glucose	↑ **LGA** -3'FSL ↑ **HGA** –Sugars*, Glucose	↑ **VD** – sugars*		
N-acetylglycoproteins (e.g *N*-acetylglutamate)	↑ **≤ 30d pp**					↑ **More excl.BM**	
SCFA (e.g. acetate, butyrate, propionate)	↑ **≥ 120d pp** – Butyrate, Propionate, Acetate	↑ **Preterm** – Butyrate	↑ **LBW** – Butyrate, Propionate	↑ **LGA** – Butyrate, Propionate		↑ **less excl.BM** Butyrate, Propionate, Acetate	
BCAA (e.g. leucine, valine)	**≥ 120d pp** – Val	↑ **Term** - Leu	↑ **HBW**-Leu, Val	↑ **HGA** - Leu, Val	↑ **VD** – Ala		
Amino acid and derivatives (e.g. tyramine, lysine, aspartate)	↑ **≥ 120d pp** – Ala, Asp, Phe, Tyr	↑ **Preterm** –Tyr ↑ **Term** – Ty, Ala, N^3-TML, Phe	↑ **LBW** – Tyr ↑ **HBW**- Ty, Ala, N^3-TML	↑ **LGA** –Ala, Tyr ↑ **HGA** – Ty, N^3-TML, Phe	↑ **VD** – Tyr, GABA ↑ **CS** – Ty	↑ **less excl. BM** Lys ↑ **more excl.BM** 3-HB	↑ **HBMI**- 4-HPA ↑ **LBMI** –Asp, 4-HPA, Gln
Methylamines & choline metabolism (e.g. Betaine)	↑ **120d pp** – Methylamine	↑ **Term**- Choline, Betaine	↑ **HBW** – Choline	↑ HGA – Choline	↑ **VD** – DMA, TMAO	↑ **more excl. BM** choline	↑ **HBMI**-TMA
Bile acids (e.g cholate and unidentified)	↑ **120d pp** – unidentified	↑ **Term** – unidentified	↑ **HBW** – unidentified				

Lipids (e.g. L1)		↑ **Preterm** -L1	↑ **LBW** –L1	↑ **LGA** – L1			
Nucleotide salvage (e.g. uracil, adenine)	↑ **≥ 120d pp** – uracil, adenine				↑ **VD** - adenine		
Other compounds (e.g. myo-inositol, 4-cresol)	↑ **120d pp** 4-cresol-glucuronide	↑ **Preterm** – glycerol			↑ **VD**- Acetone		

Table 6-2. Summary of the main urine (blue) and faecal water (black) discriminatory metabolites and their biochemical pathway from infants born term and preterm. Key: Amino Acids 3 letter key used. TMA – trimethylamine, DMA- dimethylamine, DMG- dimethylglycine, TMAO – trimethylamine-*N*-oxide, NMN – *N*-1-methylnicotinamide, *N*-AGlu - *N*-acetylglutamate, *N*-AGalS- *N*-acetylgalactosamine 6–sulphate, HomoGen - Homogentistic acid, AcetylHomoSer- Acetylhomoserine, PhoCho – Linoleoylglycerophosphocholine, 6HNA- Hydroxynicotinic Acid, N7-MeG- 7-methylguanine, HPHPA -3-hydroxy-3-(3-hydroxyphenyl) propanoate, MTFCA- 5-(2-Methylpropyl)tetrahydro-2-oxo-3-furancarboxylic acid. (*overlapped resonances from glucose, other sugars and polyols, *N*-methylated compounds and amino acids)

Class of compound/ biochemical pathway	**Discriminatory metabolites correlated to groups from multivariate statistics (OPLS, OPLS-DA)**			
	Term vs Preterm	**Term vs Preterm at birth**	**Term at birth vs Preterm at TEA**	**Term vs Preterm at 3m PP**
TCA cycle intermediates (e.g. succinate, citrate)		↑ **Term** - Citrate		
Sugars and oligosaccharides (e.g. glucose, fructose, mannitol)	↑ **Preterm** – fructose, mannitol	↑ **Preterm** – sugars*		↑ **Preterm** – fructose, mannitol
N-acetylglycoproteins intermediates and derivatives	↑ **Preterm** – N-AGalS			↑ **Term** – *N*-AGlu
BCAA (e.g. leucine, valine, α-aminoisobutyrate)		↑ **Preterm** – α-aminoisobutyrate	↑ **Preterm**- Leu	
Amino acid and derivatives (e.g. tyramine, lysine, aspartate)	↑ **Preterm** – Lys, His ↑ **Term** – HomoGen, AcetylHomoSer	↑ **Preterm** – Thr, Gln ↑ **Preterm** – Ala, Phe ↑ **Term** – Pro, HomoGen, AcetylHomoSer	↑ **Preterm** – Hyp, His ↑ **Preterm** – Ala, Lys, Ty ↑ **Term** – Pro, HomoGen, His	↑ **Preterm** - Lys

Methylamines or choline metabolism (e.g. Betaine, TMA, DMG, DMA and TMAO)	↑ **Term** – TMA, DMA, DMG, Betaine, PhoCho	↑ **Term** – TMA		↑ **Term** – DMA, DMG & Betaine
Bile acids (e.g cholate and unidentified)	↑ **Term** – unidentified, cholate	↑ **Term** – unidentified	↑ **Term** – unidentified	↑ **Preterm** – unidentified
Nutritional metabolites and Peptides, Protein degradation	↑ **Term** – polypeptides	↑ **Term** – MG, Peptides	↑ **Term** – Peptides, HPHPA, MTFCA, Tyrosol-4-sulfate, ↑ **Preterm** – polypeptide, 4- cresol	
NAD cycle, nicotinamide pathway (e.g. NMN)	↑ **Preterm** – NMN, 6HNA	↑ **Preterm** - NMN	↑ **Preterm** - NMN	↑ **Preterm** - NMN
Polyol e.g. glycerol	↑ **Preterm** – Glycerol			
Vitamins (e.g. Vit B5 and B6)	↑ Preterm – Vit B_1 ↑ **Preterm** - Vit C	↑ **Preterm**- Vit C	↑ **Preterm** – Vit B_5, Vit B_6	
Nucleontide salvage (e.g. uracil, adenine)	↑ **Term** - Pseudouridine ↑ **Preterm**- N7-MeG	↑ **Term** - Pseudouridine		↑ **Term** - Pseudouridine
Other organic acids (e.g. myo-inositol, 4-cresol)	↑ **Term** – Malonate ↑ **Preterm** – Myo-inositol	↑ **Preterm** – Myo-inositol	↑ **Term** - Creatinine	

6.3 NUTRITIONAL INFLUENCE ON INFANT DEVELOPMENT

Nutrition in early life has always been an interesting subject for studies into infant development. Breastfeeding is recommended in infants due to the complex mixture human oligosaccharides present in the milk as well as the immunological benefit for the child. Human milk oligosaccharides (HMO) are unique complex carbohydrates and are not digested by human enzymes so can reach the intestinal to promote the growth of beneficial bacteria; so act as a prebiotic. Also, HMOs purpose is to help to prevent neonatal diarrhoea and respiratory tract infections (Andreas et al 2015). Within this project, many nutritional metabolites, and energy production metabolites were evidently imperative for infant development. Changes in the excretion rates of components from the TCA cycle, amino acids, sugars and choline oxidation pathway featured in nearly all of the investigations in this thesis.

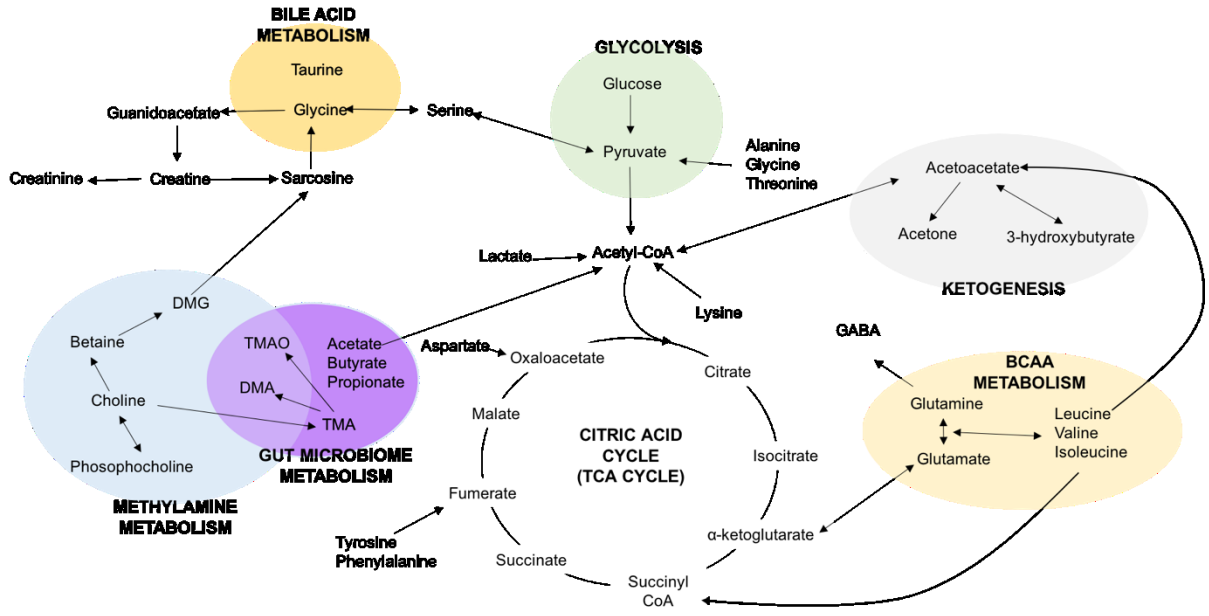


Figure 6-1: Schematic to represent metabolic pathways in developing infants

Choline and methylamine derivatives, were shown to exhibit considerable differences in all the investigations; including term versus preterm, mode of delivery, nutrition and BMI status. Trimethylamine (TMA), dimethylamine (DMA), dimethylglycine (DMG) and betaine are derived from choline, which can be obtained from the diet or endogenous sources by the breakdown of phospholipids. From the literature, betaine has been previously noted to be high at birth and increasing in concentration through the neonatal period (Trump et al 2006; Foxall et al 1995; Gu et al 2009). It is thought that at birth, betaine is elevated due to the high concentration of choline transport from the mother through the placenta to the foetus in late pregnancy. In this project, samples taken at 3 months postpartum from term born infants had elevated levels of betaine, which agrees with these reports (Foxall et al 1995; Trump et al 2006). Infant nutrition, whether breast milk or formula, is relatively high in dietary choline so this may exceed the needs of the infant and therefore betaine, a metabolic product of choline is excreted in the urine.

From the results, *N*-acetylglycoprotein fragments, sugars and choline are important for nutritional sources in infants. When stratifying the data into different lengths of time during which the infants were exclusively breast fed, they were not related to the amount of sugars excreted suggesting that this metabolite is unrelated to specific nutrition. However, choline derivatives and *N*-acetylglycoprotein fragments as well as 3-hydroxybutryate are related to infants exclusively fed breast milk for longer which agrees with the literature (Martin et al 2014).

Furthermore, once infants reach about 4-6 months of age, their nutrition changes and are typically weaned off milk onto solid foods. This change in nutrition can clearly be recognised in the faecal metabolic profile. Elevated concentration of succinate is available for the infant at this time point, suggesting a higher production of energy. Additionally, more SCFAs and 4-cresyl glucuronide are being produced at this time point compared to earlier samples which suggests a richer nutrient source for the gut microbiome.

6.4 GUT MICROBIOME INTERACTIONS

The gut microbiome plays a key role in the human health and disease, with dysbiosis causing complex diseases, including inflammatory bowel disease, celiac disease and colorectal cancer. However, this mutualistic relationship is essential for fermentation of carbohydrates and production of essential vitamins as well as inducing immune development. Host-microbe interactions and microbial co-metabolites can be recognised in all the investigations in this thesis.

SCFAs are generally regarded as beneficial to health as they prevent diarrhoea and are a nutrient source for colonic cells as well as taken up into the human body as an energy source. The most surprising results from the term vs preterm study is that there were no significantly different short chain fatty acids (SCFAs) between the two groups of infants. This is interesting as these molecules are produced by bacterial metabolism of undigestible carbohydrates and as preterm infants usually have perturbed bacterial communities this could be an expected difference between the infants. However, the NoMic study showed significant elevated concentrations of butyrate in preterm infants as well as propionate in low birth weight infants. These SCFAs are related to consuming less breast milk during the neonatal period and in high concentrations has been recognised in the potential contribution to necrotising enterocolitis (NEC) in preterm infants (Pourcyrous et al 2014; Favre et al 2002; Cristofalo et al 2013). Furthermore, SCFAs were significantly higher in infants fed formula milk, which suggests that their bacterial community were differentially dependent on nutrition. This is especially interesting as acetate was elevated in these infants, which is usually attributed to breast milk fed babies, and propionate and butyrate to formula fed infants (Parrett et al 1997).

Human breast milk contains oligosaccharides, these are undigested by the human bowel and utilised by the infant gut microbiome. The oligosaccharide, 3'fucosyllactose, was positively correlated to preterm infants implying that the gut microbiome was not fully utilising this source of nutrition, and

also supporting evidence that the preterm gut microbiome is less diverse compared to term infants (Cong et al 2016). This said however, the preterm gut microbiome is contributing to metabolic health as samples from preterm infants had elevated levels of 4-cresol.

As previously mentioned, choline and its intermediates are being utilised by the gut microbiome and are positively correlated in term born infants, vaginally delivered and exclusively fed on breast milk for longer periods. Interestingly, TMA was positively correlated to term infants at the time of birth which gives further evidence that the birth metabolic profile is influenced by the maternal metabolic profile or that there is bacterial activity in the foetus. Interestingly, vaginally delivered infants had elevated levels of bacterial co-metabolites including TMAO and GABA. Furthermore, TMA is elevated in higher BMI status infants suggesting that the gut microbiome has a role in the risk of obesity. Additionally, 4-HPA is positively correlated to infants with low and high BMI status at different time points, adding to the implication that further understanding of the composition of the gut microbiome is needed.

These microbial co-metabolites demonstrate the importance of the gut microbiome on human health especially as they are contributing to infant metabolism early in life and therefore have an impact on infant development.

6.5 KIDNEY DEVELOPMENT

At birth, renal function is sufficient for the needs of the infant to regulate electrolytes in the blood and remove waste products from the body. However, it is believed the filtration function is still developing after birth (Drukker et al 2002).

Choline metabolites including, TMAO, TMA, betaine and methylamine also function as renal osmolytes and elevation in urine and faces may be indicative of medullary damage (Foxall et al 1995). This can be been attributed to the capability of the medullary cells in the kidneys being

inefficient to take up betaine into the cell. As many of these metabolites were elevated in term infants compared to preterm infants, it suggests that renal function is not fully formed at birth and the kidneys mature further during the neonatal period. Furthermore, term infants have elevated urinary levels of pseudouridine, which has been attributed as a metabolic marker for kidney nephrotoxicity. However, preterm infants have relatively higher concentrations of 1-*N*-methylnicotiamide which is a marker of proximal tubular necrosis and renal injury (Trump et al 2006). Also, preterm infants excrete larger rates of myo-inositol in their urine which suggests that renal function is less efficient than term born infants as the kidneys are not fully absorbing all the myo-inositol before it is excreted.

Metabolites related to renal function from term and preterm infants are recognised up until 3 months of age postpartum. These kidney related metabolites found in term and preterm infant urinary metabolic profiles gives evidence that there is further development of kidney function from birth. This difference has previously been recorded. However, from this project it is evident that term and preterm infant excrete different kidney related metabolites, suggesting that the developmental trajectory are not the same depending on gestational age at birth.

6.6 STRENGTHS AND WEAKNESSES

A particular strength of this PhD thesis was the ability to standardise and explore infant collection and preparation protocols for infant profiling. This was important due to the nature of infant samples being low in volume and different analytical platforms were required for this investigation.

The NoMic study has given us the ability for pioneering exploration into the faecal metabolic profile and infant development. Human faecal metabolic profiling is still in its infancy but recently more studies are utilising this biofluid; giving further understanding into interactions among host, microbial communities and nutritional interventions. Only a handful of papers have used this biofluid for infant studies and what makes this PhD project more unique is the large participant

numbers; large participant numbers increase the strength of multivariate models and therefore the confidence in findings concluded. However, a significant limitation in this study is that matching urine samples were not available for metabolic profiling. Using urine samples gives us further understanding of which metabolites have systemically entered the body and therefore contributed to the human metabolism. This data combined with the faecal data would deepen our knowledge of metabolic profiling in the developing infant.

Unfortunately, the second study used in this thesis was actually two studies put together; term and preterm samples were obtained separately. This meant that there are many pieces of specific information missing from the metadata and a further in depth investigation into potential confounders was not possible; this included antibiotic use as well as mode of delivery. However, this study has shown the importance of further understanding the neonatal period metabolically depending on gestational age at birth to improve strategies for neonatal care of preterm infants especially in regard to the gut microbiome health and nutrition.

6.7. FUTURE WORK

Within this PhD, the two clinical projects described in chapters 4 and 5 can be taken further by future work described here. The term versus preterm at birth study would benefit by correlating metabolites from both NMR and HILIC-MS to specific microbial species. The functionality of the gut microbiome can be further explored by linking certain metabolites and microbial species which are typically found in term or preterm infant guts. This deeper investigation may open up new areas of research and understanding into the early life gut microbiome establishment and composition.

Furthermore, future work with the NoMic sample set would be to use these samples for targeted metabolic profiling, especially obtaining quantitative data of SCFA and 16S data to further explore the infant microbiome. Additionally, this data can be correlated to further investigate the functionality of the gut microbiome in early life.

REFERENCES

Aagaard, K. et al., 2014. The placenta harbors a unique microbiome. *Science translational medicine*, 6(237), p.237.

Abitbol, C.L. & Rodriguez, M.M., 2012. The long-term renal and cardiovascular consequences of prematurity. *Nature Reviews Nephrology*, 8(5), pp.265–274.

Adams, J.B. et al., 2011. Gastrointestinal flora and gastrointestinal status in children with autism--comparisons to typical children and correlation with autism severity. *BMC gastroenterology*, 11(1), p.22.

Agostoni, C. et al., 2000. Free glutamine and glutamic acid increase in human milk through a three- month lactation period. *J. Pediatr. Gastroenterol. Nutr.*, 31(5), pp.508–512.

Alberti et al., 2004. The metabolic syndrome — a new worldwide definition. *The Lancet*. 366, pp.13–16.

Altieri, L. et al., 2011. Urinary p-cresol is elevated in small children with severe autism spectrum disorder. *Biomarkers: biochemical indicators of exposure, response, and susceptibility to chemicals*, 16(3), pp.252–260.

Andreas, N.J. et al., 2016. Role of human milk oligosaccharides in Group B Streptococcus colonisation. *Clin. Translat. Immunol.*, (April), pp.1–6

Andreas, N.J., Kampmann, B. & Mehring Le-Doare, K., 2015. Human breast milk: A review on its composition and bioactivity. *Early Human Development*, 91(11), pp.629–635.

Arboleya, S. et al., 2012. Deep 16S rRNA metagenomics and quantitative PCR analyses of the premature infant fecal microbiota. *Anaerobe*, 18(3), pp.378–380.

Arboleya, S. et al., 2012. Establishment and development of intestinal microbiota in preterm neonates. *FEMS Microbiology Ecology*, 79(3), pp.763–772.

Arenz, S. et al., 2004. Breast-feeding and childhood obesity – a systematic review. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity*, 28(10), pp.1247–1256.

Arrieta, M.-C. et al., 2015. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Science Translational Medicine*, 7(307), pp.1–14.

Arslanoglu, S. et al., 2012. Amendment to 2010 Italian guidelines for the establishment and operation of a donor human milk bank. *Journal of biological regulators and homeostatic agents*, 26(3), pp.61–64.

Arslanoglu, S. et al., 2013. Donor Human Milk for Preterm Infants. *Journal of Pediatric Gastroenterology and Nutrition*, 57(4), pp.535–542.

Atzori, L. et al., 2011. 1H NMR-based metabolomics analysis of urine from preterm and term neonates. *Frontiers in Bioscience*, E3, pp.1005–1012.

Azad, M.B., et al., 2013. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ*. 185(5): pp. 385-94.

Ballard, O. & Morrow, A.L., 2013. Human Milk Composition. Nutrients and Bioactive Factors. *Pediatric Clinics of North America*, 60(1), pp.49–74.

Barrett, E., et al., 2012 gamma-Aminobutyric acid production by culturable bacteria from the human intestine. *J Appl Microbiol*,. 113(2): p. 411-7.

Bauer, S. et al., 2003. Urinary tract infection in very low birth weight preterm infants. *The Pediatric infectious disease journal*, 22(5), pp.426–30.

Baxmann, A.C. et al., 2008. Influence of muscle mass and physical activity on serum and urinary creatinine and serum cystatin C. *Clinical Journal of the American Society of Nephrology*, 3(2), pp.348–354.

Beckonert, O. et al., 2007. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc*, 2(11), pp.2692–2703.

Beecher, W.W. 2003. The Human metabolome. Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis: Its Role in Biomarker Discovery and Gene Function Analysis. Kluwer Academic Publishers.

Begley, M., Gahan, C.G.M. & Hill, C., 2005. The interaction between bacteria and bile. *FEMS Microbiology Reviews.* 29(4), pp.625–651.

Behrman, R.E. & Butler, A.S., 2007. Preterm Birth: Causes, Consequences, and Prevention. *The National Academies Press*

Bender, D.A & Mayes P.A 1996. Chapter 17. The Citric Acid Cycle: The Catabolism of Acetyl-CoA . Harper's Illustrated Biochemistry. 29th edition. McGraw-Hill Global Education Holdings, LLC

Bertino, E. et al., 2013. Benefits of donor milk in the feeding of preterm infants. *Early Human Development*, 89, pp.S3–S6.

Biasucci, G. et al., 2008. Cesarean Delivery May Affect the Early Biodiversity of Intestinal Bacteria. *The Journal of Nutrition*, 138(9), p.1796S–1800S.

Biasucci, G. et al., 2010. Mode of delivery affects the bacterial community in the newborn gut. *Early Human Development*, 86, pp.13–15.

Biro, F.M. & Wien, M., 2010. Childhood obesity and adult morbidities. *Am J Clin Nutr*, 91(1), pp.1499–1505.

Bischoff, S.C., 2011. "Gut health": a new objective in medicine? *BMC medicine*, 9(1), p.24.

Blakey, J.L., et al., 1982. Development of Gut Colonisation in Pre-term Neonates. *Journal of Medical Microbiology*. 15(4): p. 519-529.

Blencowe, H. et al., 2012. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications. *The Lancet*, 379(9832), pp.2162–2172.

Bolte, E.R., 1998. Autism and Clostridium tetani. *Medical hypotheses*, 51(2), pp.133–144.

Bouatra, S. et al., 2013. The Human Urine Metabolome. *PLoS one*, 8(9). e73076.

Boudry. G, et al., 2013. Dietary Protein Excess during Neonatal Life Alters Colonic Microbiota and Mucosal Response to Inflammatory Mediators Later in Life in Female Pigs, *J. Nutr.* 143: pp1225–1232.

Breitbart, M. et al., 2008. Viral diversity and dynamics in an infant gut. *Research in Microbiology* 159, pp. 367-373

Brennan, A-M., Murphy, B.P. & Kiely, M.E., 2016. Optimising preterm nutrition: present and future. *Proceedings of the Nutrition Society*, 75, pp.154–161.

Bronk, R. 1999. *Human metabolism* (Addison-Wesley Longman, Harlow, England).

Burke, V. et al., 2005. Breastfeeding and overweight: Longitudinal analysis in an Australian birth cohort. *Journal of Pediatrics*, 147(1), pp.56–61.

Butel, M.-J. et al., 2007. Conditions of bifidobacterial colonization in preterm infants: a prospective analysis. *Journal of pediatric gastroenterology and nutrition*, 44(5), pp.577–582.

Carlo W.A. 2016. The newborn infant. Chapter 94. Nelson Textbook of Pediatrics. 20th Edition. Elsevier Ltd.

Carroll SG et al., 1996. Lower genital tract swabs in the prediction of intrauterine infection in preterm prelabour rupture of the membranes. *Br J Obstet Gynaecol*. 103: 54–9.

Caudill, M.A., 2010. Pre- and Postnatal Health: Evidence of Increased Choline Needs. *Journal of the American Dietetic Association*, 110(8), pp.1198–1206.

Cheng, W.L. et al., 1996. Bioavailability of choline and choline esters from milk in rat pups. *Journal of Nutritional Biochemistry*, 7(8), pp.457–464.

Chiu, C.-Y. et al., 2016. Metabolomics Reveals Dynamic Metabolic Changes Associated with Age in Early Childhood. *PloS one*, 11(2), p.e0149823.

Chow, J. et al., 2014. Fecal metabolomics of healthy breast-fed versus formula-fed infants before and during in vitro batch culture fermentation. *Journal of Proteome Research*, 13(5), pp.2534–2542.

Clark, D.A., 1977. Times of first void and first stool in 500 newborns. *Pediatrics*, 60(4), pp.457–9.

Clarridge, J.E. & Alerts, C., 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin. Microbiol. Rev.*, 17(4), pp.840–862.

Claus, S.P. et al., 2008. Systemic multicompartmental effects of the gut microbiome on mouse metabolic phenotypes. *Molecular systems biology*, 4(219), p.219.

Cloarec, O. et al. 2005. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. *Anal Chem*, 77, 517-26.

Cloarec, O. et al., 2005. Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic 1 H NMR Data Sets. *Anal. Chem.*, 77(5), pp.1282–1289.

Collado, M.C. et al., 2012. Microbial ecology and host-microbiota interactions during early life stages. *Gut microbes*, 3(4), pp.352–365.

Collins, S.M. & Bercik, P., 2009. The Relationship Between Intestinal Microbiota and the Central Nervous System in Normal Gastrointestinal Function and Disease. *Gastroenterology*, 136(6), pp.2003–2014.

Cong, X. et al., 2016. Gut Microbiome Developmental Patterns in Early Life of Preterm Infants: Impacts of Feeding and Gender. *PloS One*, 11(4), p.e0152751.

Constantinou, M.A. et al., 2005. 1H NMR-based metabonomics for the diagnosis of inborn errors of metabolism in urine. *Analytica Chimica Acta*, 542(2), pp.169–177.

Cristofalo, E.A. et al., 2013. Randomized Trial of Exclusive Human Milk versus Preterm Formula Diets in Extremely Premature Infants. *The Journal of Pediatrics*, 163(6), pp.1592–1595.

Darlow, B.A. et al., 2005. Vitamin C supplementation in very preterm infants: a randomised controlled trial. *Archives of disease in childhood. Fetal and neonatal edition*, 90(2), pp.F117-122.

De Hoffman, E., Charette, J. & Stroobant, V. 1996. Mass spectrometry: Principles and applications, New York, Wiley.

De Jong, F. et al., 2012. Systematic review and meta-analysis of preterm birth and later systolic blood pressure. *Hypertension*, 59(2), pp.226–234.

Deda, O. et al., 2015. An overview of fecal sample preparation for global metabolic profiling. *Journal of Pharmaceutical and Biomedical Analysis*, 113, pp.137–150.

Despres, J.-P. & Lemieux, I., 2006. Abdominal obesity and metabolic syndrome. Nature, 444(7121), pp.881–887.

Dessì, A. et al., 2014. Investigation of the ^{1} H-NMR based urine metabolomic profiles of IUGR, LGA and AGA newborns on the first day of life. *The Journal of Maternal-Fetal & Neonatal Medicine*, 27(sup2), pp.13–19.

Dettmer, K., Aronov, P. A. and Hammock, B. D. 2007. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* 26: 51–78.

Diaz, S.O. et al., 2016. Newborn Urinary Metabolic Signatures of Prematurity and Other Disorders: A Case Control Study. *Journal of Proteome Research*, 15(1), pp.311–325.

Dieterle, F., et al. 2011. NMR and MS methods for metabonomics. *Methods Mol Biol*. 691: p. 385-415.

Dominguez-Bello, M.G. et al., 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26), pp.11971–5.

Dona, A.C. et al., 2014. Precision High-Throughput Proton NMR Spectroscopy of Human Urine, Serum, and Plasma for Large-Scale Metabolic Phenotyping. *Anal. Chem.*, 86, pp.9887–9894.

Dong, Q. et al., 2011. Immunology and Microbiology Diversity of Bacteria at Healthy Human Conjunctiva *Immunology and Microbiology*. 52(8), pp.5408–5413.

Druet, C. et al., 2012. Prediction of childhood obesity by infancy weight gain: An individual-level meta-analysis. *Paediatric and Perinatal Epidemiology*, 26(1), pp.19–26.

Drukker, A. 2002. The adverse renal effects of prostaglandin-synthesis inhibition in the fetus and the newborn. *Paediatrics & Child Health*, 7(8), pp.538–543.

Duarte, I.F. et al., 2009. Identification of metabolites in human hepatic bile using 800 MHz 1H NMR spectroscopy, HPLC-NMR/MS and UPLC-MS. *Molecular bioSystems*, 5(2), pp.180–190.

Duerkop, B.A. & Hooper, L. V, 2013. Resident viruses and their interactions with the immune system. *Nat Immunol*, 14(7), pp.654–659.

Dumas, M.E*., et al.* 2006*.* Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc Natl Acad Sci U S A* 103, pp.12511-12516

Eckburg et al 2005. Diversity of the Human Intestinal Microbial Flora. *Science*. 308. pp1635–1638.

Edmond and Bahl., 2006. Optimal feeding of low-birth-weight infants. Technical review. WHO Library.

Edwards, C.A. & Parrett, A., 2002. Intestinal flora during the first months of life: new perspectives. *British Journal of Nutrition*, 88, pp.S11–S18.

Eggesbo, M. et al., 2011. Development of gut microbiota in infants not exposed to medical interventions. *Apmis*, 119(1), pp.17–35.

Fallani, M. et al., 2010. Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. *Journal of pediatric gastroenterology and nutrition*, 51(1), pp.77–84.

Fanaro, S., 2013. Early Human Development Feeding intolerance in the preterm infant. *Early Human Development,* 89, pp.S13–S20.

Fanos, V. et al., 2012. Clinical application of metabolomics in neonatology. *Journal of Maternal-Fetal and Neonatal Medicine*, 25(S1), pp.104–109.

Fanos, V. et al., 2012. Urinary Metabolomics in Newborns and Infants. *Advances in Clinical Chemistry*, 58, pp.194-218.

Fanos, V. et al., 2013. Metabolomics application in maternal-fetal medicine. *BioMed Research International*, 2013.

Fanos, V. et al., 2013. Metabolomics in neonatology: Fact or fiction? *Seminars in Fetal and Neonatal Medicine*, 18(1), pp.3–12.

Favre, A. et al., 2002. Diet, Length of Gestation, and Fecal Short Chain Fatty Acids in Healthy Premature Neonates. *Journal of Parenteral and Enteral Nutrition* 26. pp51–56

Finegold, S.M., 2008. Therapy and epidemiology of autism-clostridial spores as key elements. *Medical Hypotheses*, 70(3), pp.508–511.

Flint, H.J. et al., 2012. The role of the gut microbiota in nutrition and health. *Nature Reviews Gastroenterology and Hepatology*, 9(10), pp.577–589.

Ford, J.E., et al., 1983. Comparison of the B vitamin composition of milk from mothers of preterm and term babies. *Archives of Disease in Childhood.* 58(5): p. 367-372.

Fouhy, F. et al., 2012. High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. *Antimicrobial Agents and Chemotherapy*, 56(11), pp.5811–5820.

Fox, G.E. et al., 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization (comparative oligonucleotide cataloging/phylogeny/molecular evolution). *Evolution*, 74(10), pp.4537–4541.

Foxall, P.J. et al., 1995. Analysis of fetal and neonatal urine using proton nuclear magnetic resonance spectroscopy. *Archives of disease in childhood. Fetal and neonatal edition*, 73(3), pp.F153-7.

Frank, D.N. & Pace, N.R., 2008. Gastrointestinal microbiology enters the metagenomics era. *Current opinion in gastroenterology*, 24(1), pp.4–10.

Frey, H.A. & Klebanoff, M.A., 2016. The epidemiology, etiology, and costs of preterm birth. *Seminars in Fetal & Neonatal Medicine*, 21(2), pp.68–73

Gephart, M. et al., 2013. Necrotizing Enterocolitis Risk: State of the Science, *Adv Neonatal Care.* 12(2), pp.77–89.

Geuking, M.B. et al., 2014. The interplay between the gut microbiota and the immune system. *Gut Microbes*, 5(3).

Gill et al. 2006. Metagenomic Analysis of the Human Distal Gut Microbiome. *Science*. 312(5778): pp.1355–1359.

Goldenberg, R.L. et al., 2008. Epidemiology and causes of preterm birth. *The Lancet*, 371, pp.75–84.

Goldenberg, R.L., J.C. Hauth, and W.W. Andrews, 2000. Intrauterine infection and preterm delivery. *N Engl J Med,*342(20): p. 1500-7.

Gomez-Llorente, C. et al., 2013. Three main factors define changes in fecal microbiota associated with feeding modality in infants. *Journal of pediatric gastroenterology and nutrition*, 57(4), pp.461–6.

Gondalia, S.V. et al., 2010. Faecal microbiota of individuals with autism spectrum disorder. *E-Journal of Applied Psychology*, 6(2), pp.24–29.

Goodpaster, A.M., Ramadas, E.H. & Kennedy, M.A., 2011. Potential Effect of Diaper and Cotton Ball Contamination on NMR- and LC/MS-Based Metabonomics Studies of Urine from Newborn Babies. *Anal. Chem.*, 83, pp.896–902.

Grantham-McGregor, S.M. et al., 1998. The development of low birth weight term infants and the effects of the environment in Northeast Brazil. *Journal of Pediatrics*, 132(4), pp.661–666.

Groer, M.W. et al., 2014. Development of the preterm infant gut microbiome: a research priority. *Microbiome*, 2(1), p.38.

Gu. H et al., 2009. ^{1}H NMR metabolomics study of age profiling in children. *NMR Biomed*. 22(8): 826–833.

Guarner, F. & Malagelada, J.R., 2003. Gut flora in health and disease. *Lancet* 361, 512-519

Gupta, S. et al., 2011. Metagenome of the gut of a malnourished child. *Gut Pathogens*, 3, pp.7–16.

Gutcher, R. & Farrell, M., 1991. Intravenous infusion of lipid for the prevention fatty acid deficiency in premature infants13 of essential syndrome. *Am J Clin Nutr*, 54, pp.1024–1028

Hague, A., Butt, A.J. & Paraskeva, C., 1996. The role of butyrate in human colonic epithelial cells: an energy source or inducer of differentiation and apoptosis? *The Proceedings of the Nutrition Society*, 55(3), pp.937–43.

Hällström, M. et al., 2004. Effects of mode of delivery and necrotising enterocolitis on the intestinal microflora in preterm infants. *European Journal of Clinical Microbiology and Infectious Diseases*, 23(6), pp.463–470.

Handelsman, J., 2004. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and molecular biology reviews*, 68(4), pp.669–685.

Handley, S.A. et al., 2012. Pathogenic Simian Immunodeficiency Virus Infection Is Associated with Expansion of the Enteric Virome. *Cell*. 151, pp.253–266.

Hanna, M.H. et al., 2013. Urinary metabolomic markers of aminoglycoside nephrotoxicity in newborn rats. *Pediatr REs*, 73(5), pp.585–591.

Hao, H. et al., 2015. Metabolic products in urine of preterm infants characterized via gas chromatography-mass spectrometry. *International Journal of Clinical and Experimental Medicine*, 8(9), pp.16454–16462.

Harmsen, H.J. et al., 2000. Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. *Journal of pediatric gastroenterology and nutrition*, 30(1), pp.61–67.

Hay, W.W. & Hendrickson, K.C., 2016. Preterm formula use in the preterm very low birth weight infant. *Seminars in Fetal & Neonatal Medicine*, (Article in press).

Ho, M.-Y. & Yen, Y.-H., 2015. Trend of Nutritional Support in Preterm Infants. *Pediatrics and Neonatology*, (Article in press).

Holmes-McNary M.Q. et al. 1996. Choline and choline esters in human and rat milk and in infant formulas. *Am J Cli Nutr*. 64. pp572-6.

Holmes, E. et al., 2011. Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends in Microbiology*, 19, pp.349–359.

Holmes, H.C., Snodgrass, G.J. & Iles, R. a, 2000. Changes in the choline content of human breast milk in the first 3 weeks after birth. *European journal of pediatrics*, 159(3), pp.198–204.

Hooper, L. V & Gordon, J.I., 2001. Commensal host-bacterial relationships in the gut. *Science (New York, N.Y.)*, 292(5519), pp.1115–1118.

Hooper, L. V., Littman, D.R. & Macpherson, A.J., 2012. Interactions Between the Microbiota and the Immune System. *Science*, 336(6086), pp.1268–1273.

Hooper, L.V., 2004 Bacterial contributions to mammalian gut development. *Trends in Microbiology*,. 12(3): p. 129-134.

Hopwood, J.J. & Elliott, H., 1985. Urinary excretion of sulphated N-acetylhexosamines in patients with various mucopolysaccharidoses. *The Biochemical journal*, 229(3), pp.579–586.

Horta, B.L. et al., 2013. Birth by Caesarean Section and Prevalence of Risk Factors for Non-Communicable Diseases in Young Adults: A Birth Cohort Study. *PLoS ONE*, 8(9), pp.1–9.

Hoult, D. & Richards, R., 1976. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *Journal of Magnetic Resonance*, 24(1), pp.71–85.

Houten, S. M., Watanabe, M., & Auwerx, J. 2006. Endocrine functions of bile acids. *The EMBO Journal,* 25(7), pp.1419–1425.

Howie, P. et al., 1990. Protective effect of breast feeding against infection. *BMJ*. 300: pp.11-6

Howson, C.P. et al., 2013. Born too soon: preterm birth matters. *Reproductive health*, 10(S1), p.S1.

Hyde et al., 2010. The effect of preterm delivery on the urinary metabolome. *The Neonatal Society Abstracts.* 2010 Summer Meeting.

Ilcol, Y.O. et al., 2005. Choline status in newborns, infants, children, breast-feeding women, breast-fed infants and human breast milk. *Journal of Nutritional Biochemistry*, 16, pp.489–499

Inagaki, T. et al., 2006. Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor. *PNAS*. 103(10) pp 3920–3925.

Irving, R.J. et al., 2000. Adult cardiovascular risk factors in premature babies. *Lancet*, 355, pp.2135–2136.

Ivorra, C. et al., 2012. Metabolomic profiling in blood from umbilical cords of low birth weight newborns. *Journal of Translational Medicine*, 10(1), p.142.

Jackson, F. et al., 2016. Development of a Pipeline for Exploratory Metabolic Profiling of Infant Urine. *Journal of Proteome Research*, 15(9), pp.3432–3440

Jacobs, D.M., et al., 2008. (1)H NMR metabolite profiling of feces as a tool to assess the impact of nutrition on the human microbiome. *NMR Biomed*. 21(6): pp. 615-26.

Jacquot, A., et al., 2011. Dynamics and Clinical Evolution of Bacterial Gut Microflora in Extremely Premature Patients. *The Journal of Pediatrics.* 158(3): pp. 390-396.

Jefferies, A.L., 2016. Going home: Facilitating discharge of the preterm infant. *Paediatr Child Health*, 19(1), pp.31–36.

Jernberg, C, Löfmark, S., Edlund, C. Jansson, J.K., 2010. Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology*. vol. 156(11) pp. 3216-3223

Jess, T., 2015. Microbiota, Antibiotics, and Obesity. *The New England Journal of Medicine*, 371(26), pp.2014–2016.

Jimenez, E. et al., 2008. Is meconium from healthy newborns actually sterile? *Research in Microbiology*, 159(3), pp.187–193.

Johansson, S. et al., 2005. Risk of high blood pressure among young men increases with the degree of immaturity at birth. *Circulation*, 112(22), pp.3430–3436.

Johnson, S. et al., 2010. Autism Spectrum Disorders in Extremely Preterm Children. *Journal of Pediatrics*, 156(4), p.525–531.

Jones, E. & Spencer, S.A., 2007. Optimising the provision of human milk for preterm infants. *Archives of Disease in Childhood Fetal and Neonatal Edition*, 92(4), pp.F236–F238.

Jumpertz, R. et al., 2011. Energy-balance studies reveal associations between gut microbes, caloric load , and nutrient absorption in humans. *Am J Clin Nutr*. 94, pp.58–65.

Jurtshuk, P., 1996. Chapter 4: Bacterial Metabolism. Baron S, editor. Medical Microbiology. 4th edition. University of Texas Medical Branch at Galveston.

Kau, A.L., et al., 2011. Human nutrition, the gut microbiome, and immune system: envisioning the future. *Nature*. 474(7351): p. 327-336.

Keeler, J. Understanding NMR Spectroscopy. (University of Cambridge, Department of Chemistry, 2002).

Kelly, D. & Coutts, a G., 2000. Early nutrition and the development of immune function in the neonate. *The Proceedings of the Nutrition Society*, 59, pp.177–185.

Kim et al 2004. Cloning and Characterization of the Bile Salt Hydrolase Genes (bsh) from Bifidobacterium bifidum Strains. *Applied and environmental microbiology*, 70(9) p. 5603–5612

Knol, J. et al., 2005. Colon microflora in infants fed formula with galacto- and fructo-oligosaccharides: more like breast-fed infants. *Journal of Pediatric Gastroenterology and Nutrition*, 40(1), pp.36–42.

Koenig, J.E. et al., 2011. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl, pp.4578–4585.

Koletzko, B. et al., 2012. Early nutrition programming of long-term health. *Proceedings of the Nutrition Society*, 71(3), pp.371–378.

Koleva, P.T., Bridgman, S.L. & Kozyrskyj, A.L., 2015. The infant gut microbiome: Evidence for obesity risk and dietary intervention. *Nutrients*, 7(4), pp.2237–2260.

Kotecha, S., 2000. Lung growth: implications for the newborn infant. *Archives of disease in childhood. Fetal and neonatal edition*, 82(1), pp.F69–F74.

Kull, I., Almqvist, C. & Lilja, G., 1996. Breast-feeding reduces the risk of asthma during the first 4 years of life. *J Allergy Clin Immunol.* 114:4. pp755-60

Kumps, A., Duez, P. & Mardens, Y., 2002. Metabolic, nutritional, iatrogenic, and artifactual sources of urinary organic acids: A comprehensive table. *Clinical Chemistry*, 48(5), pp.708–717.

Kurokawa K, et al. 2007. Comparative Metagenomics Revealed Commonly Enriched Gene Sets in Human Gut Microbiomes. *DNA Research* 14, pp169–181

La Rosa, P.S. et al., 2014. Patterned progression of bacterial populations in the premature infant gut. *Proceedings of the National Academy of Sciences*, 111(48), pp.17336–17336.

Lager S, and Powell TL. 2012. Regulation of nutrient transport across the placenta. *J Pregnancy*. Article ID 179827

Lapillonne, A., et al., 2013. Lipid needs of preterm infants: updated recommendations. *J Pediatr*.

162(3 Suppl): p. S37-47.

Le Chatelier, E. Le et al., 2013. Richness of human gut microbiome correlates with metabolic markers. Nature. 500, pp.541

Le Gall, G. et al., 2011. Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome. *Journal of Proteome Research*, 10(9), pp.4208–4218.

Lee, A. & Gemmell, E., 1972. Changes in the Mouse Intestinal Microflora During Weaning: Role of Volatile Fatty Acids. *Infection and Immunity*, 5(1), pp.1–7.

Leser and Molbak. 2009. Better living through microbial action: the benefits of the mammalian gastrointestinal microbiota of the host. *Environmental Microbiology*. 11(9), pp.2194–2206

Ley, R. et al., 2006. Microbial ecology: human gut microbes associated with obesity. *Nature*, 444, pp.1022–3.

Ley, R.E. et al., 2005. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31), pp.11070–5.

Li, H. & Cao, Y., 2010. Lactic acid bacterial cell factories for gamma-aminobutyric acid. *Amino Acids*, 39(5), pp.1107–1116.

Li, J. V et al., 2011. Metabolic surgery profoundly influences gut microbial-host metabolic cross-talk. *Gut*, 60(9), pp.1214–23.

Li, J. V. et al., 2011. Experimental bariatric surgery in rats generates a cytotoxic chemical environment in the gut contents. *Frontiers in Microbiology*, 2, pp.1–9.

Lifschitz, C., 2015. Early Life Factors Influencing the Risk of Obesity. *Pediatric gastroenterology, hepatology & nutrition*, 18(4), pp.217–23.

Lloyd, A J. et al., 2011. Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *Am J Clin Nutr*. 94. Pp. 981-91

Louis, P., Hold, G.L. & Flint, H.J., 2014. The gut microbiota, bacterial metabolites and colorectal cancer. *Nature Reviews Microbiology*, 12(10), pp.661–672.

Mahan L.K., 2011. Krause's Food & the Nutrition Care Process. 13th edition. Saunders.

Maitre, L. et al., 2014. Urinary metabolic profiles in early pregnancy are associated with preterm birth and fetal growth restriction in the Rhea mother-child cohort study. *BMC medicine*, 12, p.110.

Malina et al., 2004. Growth, Maturation, and Physical Activity - Robert M. Malina, Claude Bouchard, Oded Bar. 2nd Edition. Human Kinetics.

Marchesi, J.R. et al., 2007. Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. *Journal of Proteome Research*, 6(2), pp.546–551.

Marcobal, A. et al., 2012. Tyramine and Phenylethylamine Biosynthesis by Food Bacteria. *Critical Reviews in Food Science and Nutrition*, 52(5), pp.448–467.

Marcobal, A. et al., 2013. A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *The ISME journal*, 7(10), pp.1933–1943.

Mardis, E.R., 2008. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics*, 9(1), pp.387–402.

Margulies, M. et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437(7057), pp.376–380.

Martin J et al., 2013. Births: Final data for 2013. National vital statistics reports: From the centers for disease control and prevention, national center for health statistics. *National Vital Statistics System.* Vol.64, p.1-65

Martin, F.-P.J. et al., 2014. Impact of breast-feeding and high- and low-protein formula on the metabolism and growth of infants from overweight and obese mothers. *Pediatric research*, 75(4), pp.535–543.

Martin, F.J. et al., 2010. Metabonomics research articles. *Journal of Proteome Research*, pp.5284–5295.

Martín, R. et al., 2003. Human milk is a source of lactic acid bacteria for the infant gut. *Journal of Pediatrics*, 143(6), pp.754–758.

Matamoros, S. et al., 2013. Development of intestinal microbiota in infants and its impact on health. *Trends in Microbiology*, 21, pp.167–173.

Mei, J. V et al., 2001. Innovative Non- or Minimally-Invasive Technologies for Monitoring Health and Nutritional Status in Mothers and Young Children Use of Filter Paper for the Collection and Analysis of Human Whole Blood Specimens 1. *The Journal of Nutrition*, 131, pp.1631–1636.

Meyers, J.M. et al., 2013. Potential influence of total parenteral nutrition on body composition at discharge in preterm infants. *The Journal of Maternal-Fetal & Neonatal Medicine*, 26(15), pp.1548–1553.

Michopoulos, F. et al., 2009. UPLC MS based analysis of human plasma for metabonomics using solvent precipitation or solid phase extraction. *J.Proteome Res.*, 8(4), pp.2114–2121.

Miller, T.L. & Wolin, M.J., 1996. Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora. *Applied and Environmental Microbiology*, 62(5), pp.1589–1592.

Moco, S. et al., 2013. Metabolomics perspectives in pediatric research. *Pediatric research*, 73(4), pp.570–576.

Moltu, S.J. et al., 2014. Urinary metabolite profiles in premature infants show early postnatal metabolic adaptation and maturation. *Nutrients*, 6(5), pp.1913–1930.

Monleon, D., et al., 2009. Metabolite profiling of fecal water extracts from human colorectal cancer. *NMR Biomed*. 22(3): pp. 342-8.

Monteiro, P. O. A. & Victoria, C. G. 2005. Rapid growth in infancy and childhood and obesity in later life – a systematic review. *Obesity Reviews*, 6: pp143–154.

Moster, D., Lie, R.T. & Markestad, T., 2008. Long-term medical and social consequences of preterm birth. *The New England journal of medicine*, 359(3), pp.262–73.

Moutquin, J.M., 2003. Classification and heterogeneity of preterm birth. *BJOG: An International Journal of Obstetrics and Gynaecology*, 110, pp.30–33.

Mueller, N.T. et al., 2015. *The infant microbiome development: mom matters*. HHS Public Access, 21(2), pp.109–117.

Murphy, G.M, and Signer E. 1974. Progress Report: Bile acid metabolism in infants and children. *Gut*. 15: 151-163

Musacchio, T. et al., 2009. 1H NMR detection of mobile lipids as a marker for apoptosis: The case of anticancer drug-loaded liposomes and polymeric micelles. *Molecular Pharmaceutics*, 6(6), pp.1876–1882.

Musilova RV, Vlkova E, and Bunesova V. 2014. Beneficial effects of human milk oligosaccharides on gut microbiota. *Beneficial Microbes* 2014 5:3, 273-283

Musso, G., Gambino, R. & Cassader, M., 2010. Obesity, diabetes, and gut microbiota: The hygiene hypothesis expanded? *Diabetes Care*, 33(10), pp.2277–2284.

Musso, G., Gambino, R. & Cassader, M., 2011. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. *Annual review of medicine*, 62, pp.361–380.

Nava, G.M. et al., 2012. Hydrogenotrophic microbiota distinguish native Africans from African and European Americans. *Environmental Microbiology Reports*. 4, pp.307–315.

Neu, J., 2007. Gastrointestinal development and meeting the nutritional needs of premature infants. *American Journal of Clinical Nutrition*, 85(2).

Nicholson, J.K. et al., 2012. Host-Gut Microbiota Metabolic Interactions. *Science*, 336, 1262-1267. *Science*, 336, pp.1262–1267.

Nicholson, J.K. et al., 2012. Metabolic phenotyping in clinical and surgical environments. *Nature*, 491, pp.384–392.

Nicholson, J.K., Holmes, E. & Wilson, I.D., 2005. Gut microorganisms, mammalian metabolism and personalized health care. *Nature reviews. Microbiology*, 3(5), pp.431–438.

Nicholson, J.K., Lindon, J.C. & Holmes, E., 1999. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29(11), pp.1181–1189.

Nkadi P.O, Merritt T.A, & Pillers DM. 2009. An Overview of Pulmonary Surfactant in the Neonate: Genetics, Metabolism, and the Role of Surfactant in Health and Disease. *Mol Genet Metab*. 97(2): 95–101.

Norman, J. & Greer, I., 2005. Preterm Labour Managing Risk in Clinical Practice. *Cambridge University Press*

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

O'Sullivan, A. et al., 2013. Metabolomic phenotyping validates the infant rhesus monkey as a model of human infant metabolism. *Journal of pediatric gastroenterology and nutrition*, 56(4), pp.355–63.

ONS, 2015. Statistical Bulletin. Birth characteristics in England and Wales: 2014. pp1-10.

Ottman, N. et al., 2012. The function of our microbiota: who is out there and what do they do? *Front Cell Infect Microbiol*, 2(August), p.104.

Ozarda, Y., Cansev, M. & Ulus, I.H., 2014. Breast Milk Choline Contents Are Associated with Inflammatory Status of Breastfeeding Women. *Journal of Human Lactation*, 30(2), pp.161–166.

Palmer, C. et al., 2007. Development of the human infant intestinal microbiota. *PLoS Biology*, 5(7), pp.1556–1573.

Pande, G. & Akoh, C.C., 2016. Enzymatic Synthesis of Tyrosol-Based Phenolipids: Characterization and Effect of Alkyl Chain Unsaturation on the Antioxidant Activities in Bulk Oil and Oil-in-Water Emulsion. *JAOCS, Journal of the American Oil Chemists' Society*, 93(3), pp.329–337.

Parkinson, J.R.C. et al., 2013. Preterm birth and the metabolic syndrome in adult life: a systematic review and meta-analysis. *Pediatrics*, 131(4), pp.e1240-63.

Parks, D.H. et al., 2014. STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics*, 30(21), pp.3123–3124.

Parm, Ü. et al., 2011. Early Human Development Risk factors associated with gut and nasopharyngeal colonization by common Gram-negative species and yeasts in neonatal intensive care units patients. *Early Immune Development*, 87, pp.391–399.

Parracho, H.M.R.T. et al., 2005. Differences between the gut microflora of children with autistic spectrum disorders and that of healthy children. *Journal of Medical Microbiology*, 54(10), pp.987–991.

Parrett, A.M. & Edwards, C.A., 1997. In vitro fermentation of carbohydrate by breast fed and formula fed infants. *Archives of disease in childhood*, 76(3), pp.249–53.

Patel, P. & Bhatia, J., 2016. Total parenteral nutrition for the very low birth weight infant. *Seminars in Fetal & Neonatal Medicine*, (Article in press).

Pearl, P.L. et al., 2006. Inherited disorders of GABA metabolism. *Future Neurol*, 1(5), pp.631–636.

Penders, J. et al., 2006. Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy. *Pediatrics*, 118(2), pp.511–521.

Perng, W. et al., 2014. Metabolomic profiles and childhood obesity. *Obesity (Silver Spring, Md.)*, 22(12), pp.2570–2578.

Pessione, E., 2012. Lactic acid bacteria contribution to gut microbiota complexity: lights and shadows. *Frontiers in Cellular and Infection Microbiology*, 2, pp.1–15.

Pettersson, E., Lundeberg, J. & Ahmadian, A., 2008. Generations of sequencing technologies. *Genomics*, 93, pp.105–111.

Pettersson, J. et al., 2008. NMR metabolomic analysis of fecal water from subjects on a vegetarian diet. *Biological & pharmaceutical bulletin*, 31(6), pp.1192–8.

Pitt, J.J., 2009. Principles and Applications of Liquid Chromatography - Mass Spectrometry in Clinical Biochemistry. *Clin Biochem Rev*, 30(1), pp.19–34.

Pitt, J.J., 2010. Newborn screening. *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists*, 31(2), pp.57–68.

Pourcyrous, M. et al., 2014. Fecal short-chain fatty acids of very-low-birth-weight preterm infants fed expressed breast milk or formula. *J Pediatr Gastroenterol Nutr.*, 59(6), pp.725–731.

Praticò, G., et al. 2013. Exploring human breast milk composition by NMR-based metabolomics. *Natural Product Research*; 37–41.

Quigley, M.A. et al., 2016. Exclusive breastfeeding duration and infant infection. *European Journal of Clinical Nutrition*. 70: pp.1420–1427.

Radmacher, P.G. & Adamkin, D.H., 2016. Fortification of human milk for preterm infants. *Seminars in Fetal & Neonatal Medicine*, (Article in press).

Raman, M. et al., 2013. Fecal microbiome and volatile organic compound metabolome in obese humans with nonalcoholic fatty liver disease. *Clinical Gastroenterology and Hepatology*, 11(7), pp.868–875.

Ramsay, J.O. 1988. Monotone Regression Splines in Action. *Statistical Science*. 3(4) pp425-461

Reaven, G.M., 1988. Role of Insulin Resistance in Human. *Diabetes*. 37, pp.1595–1607.

Ridlon, J.M., Kang, D.-J. & Hylemon, P.B., 2006. Bile salt biotransformations by human intestinal bacteria. *Journal of lipid research*, 47(2), pp.241–259.

Roager, H.M. et al., 2016. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nature Microbiology*, 1, p.16093.

Rodriguez, J.M. et al., 2015. The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis*, 26, p.26050.

Rodwell, V. 2012. Chpt 28. Harper's Illustrated Biochemistry. 29th edition. McGraw-Hill Global Education Holdings, LLC

Roger, L.C. & McCartney, A.L., 2010. Longitudinal investigation of the faecal microbiota of healthy full-term infants using fluorescence in situ hybridization and denaturing gradient gel electrophoresis. *Microbiology*, 156(11), pp.3317–3328.

Romano, K.A. et al., 2015. Intestinal Microbiota Composition Modulates Choline Bioavailability. *mBio*, 6(2), pp.1–8.

Rotimi, V.O., Olowe, S.A. & Ahmed, I., 1985. The development of bacterial flora of premature neonates. *The Journal of hygiene*, 94, pp.309–318.

Russell, W.R. et al., 2011. High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *American Journal of Clinical Nutrition*, 93(5), pp.1062–1072.

Rutayisire, E. et al., 2016. The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. *BMC Gastroenterology*, 16(1), pp.86.

Salminen, S.J., Gueimonde, M. & Isolauri, E., 2005. Symposium: Innate Immunity and Human Milk Probiotics That Modify Disease Risk 1. *Journal of Nutrition*, pp.1294–1298.

Sander, G. et al., 1986. Ribonucleic acid turnover in man: RNA catabolites in urine as measure for the metabolism of each of the three major species of RNA. *Clinical science*, 71(4), pp.367–374.

Sanderson, I.R, & Naik, S. 2000. Dietary regulation of intestinal gene expression. *Annu. Rev. Nutr*. 20. Pp311-38

Sanger, F., Nicklen, S. & Coulson, A.R., 1977. Biochemistry DNA sequencing with chain-terminating inhibitors (DNA polymerase/nucleotide sequences/bacteriophage 4X174). *Proc. Natl. Acad. Sci. USA*, 74(12), pp.5463–5467.

Sarafian, M.H. et al., 2015. Bile Acid Profiling and Quantification in Biofluids Using Ultra-Performance Liquid Chromatography Tandem Mass Spectrometry. *Analytical Chemistry*, 87(19), pp.9662–9670.

Saric, J. et al., 2008. Species variation in the fecal metabolome gives insight into differential gastrointestinal function. *Journal of Proteome Research*, 7(1), pp.352–360.

Saude, E., et al., Variation of metabolites in normal human urine. Metabolomics, 2007. 3(4): p. 439-451.

Schanler, R.J., Lau, C., Hurst, N.M. & O'Brian Smith, E., 2005. Randomized Trial of Donor Human Milk Versus Preterm Formula as Substitutes for Mothers' Own Milk in the Feeding of Extremely Premature Infants. *Paediatrics*, 116(2), pp.400–406

Schell-Feith, E.A, et al. 2000. Etiology of nephrocalcinosis in preterm neonates : Association of nutritional intake and urinary parameters. *Kidney International* 58. pp.2102–2110.

Schloss, P.D. et al., 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23), pp.7537–7541.

Schloss, P.D. et al., 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23), pp.7537–7541.

Schulpis, K.H. et al., 2008. The effect of the mode of delivery on maternal-neonatal interleukin-6, biogenic amine and their precursor amino acid concentrations. *Clinical Chemistry and Laboratory Medicine*, 46(11), pp.1624–1630.

Schutze, G.E., Willoughby, R.E. & Committee on Infectious Diseases American Academy of Pediatrics, 2013. Clostridium difficile Infection in Infants and Children. *Pediatrics*, 131(1), pp.196–200.

Servillo, L. et al., 2014. Where does N(ε)-trimethyllysine for the carnitine biosynthesis in mammals come from? *PloS one*, 9(1), p.e84589.

Shattock, P. & Whiteley, P., 2002. Biochemical aspects in autism spectrum disorders: updating the opioid-excess theory and presenting new opportunities for biomedical intervention. *Expert opinion on therapeutic targets*, 6(2), pp.175–83.

Simmons, L.E. et al., 2010. Preventing Preterm Birth and Neonatal Mortality: Exploring the Epidemiology, Causes, and Interventions. *Seminars in Perinatology*, 34(6), pp.408–415.

Singhal, A. et al., 2001. Preterm birth, vascular function, and risk factors for atherosclerosis. *Lancet*, 358, pp.1159–1560.

Sleeth, M.L., Thompson, E.L., Ford, H.E., Zac-Varghese, S.E.K. & Frost, G., 2010. Free fatty acid receptor 2 and nutrient sensing: a proposed role for fibre, fermentable carbohydrates and short-chain fatty acids in appetite regulation', *Nutrition Research Reviews*, 23(1), pp. 135–145

Smilowitz, J. et al., 2014. Breast milk oligosaccharides: structure-function relationships in the neonate. *Annu Rev Nutr*, 34, pp.143–169.

Solis, G. et al., 2010. Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut. *Anaerobe*, 16(3), pp.307–310.

Song, S.J., Dominguez-Bello, M.G. & Knight, R., 2013. How delivery mode and feeding can shape the bacterial community in the infant gut. *Cmaj*, 185(5), pp.373–374.

Spagou, K. et al., 2010. Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. *Journal of Separation Science*, 33(6–7), pp.716–727.

Stark, P.L. & Lee, A, 1982. The bacterial colonization of the large bowel of pre-term low birth weight neonates. *The Journal of hygiene*, 89(1), pp.59–67.

Stark, P.L. & Lee, A., 1982. The Microbial Ecology of the Large Bowel of Breast-Fed and Formula-Fed Infants During the First Year of Life. *J. Med. Microbiol.*, 15, pp.189–203.

Stephens, B.E. et al., 2012. Screening for autism spectrum disorders in extremely preterm infants. *J Dev Behav Pediatr*, 33(7), pp.535–541.

Sudo N, Chida Y, Aiba Y, et al. 2004. Postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in mice. *J Physiol*. 558. pp263–275.

Sumner, L. W., et al. 2007. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 3(3), 211–221.

Synderman, S.E. 1963. The histidine requirement of the infant. *Pediatrics.* 31. pp786

Taft, DH. et al., 2016 Abstract: Differences in the Infant Gut Microbiota Related to the Fatty Acid Composition of Human Milk: Results from the GEHM Cohort. *The FASEB Journal.* vol. 30 no. 1 Supplement 406.7

Tanaka, S. et al., 2009. Influence of antibiotic exposure in the early postnatal period on the development of intestinal microbiota. *FEMS Immunology and Medical Microbiology*, 56(1), pp.80–87.

Tannock, G.W. et al., 2013. Comparison of the compositions of the stool microbiotas of infants fed goat milk formula, cow milk-based formula, or breast milk. *Applied and Environmental Microbiology*, 79(9), pp.3040–3048.

Tapp, H.S. & Kemsley, E.K., 2009. Notes on the practical utility of OPLS. *Trends in Analytical Chemistry*, 28, pp.1322–1327.

Tea I, Le Gall G, Küster A, et al. 2012. 1H-NMR-based metabolic profiling of maternal and umbilical cord blood indicates altered materno-foetal nutrient exchange in preterm infants. *PLoS One*. 7(1).

the adult phenotype of the preterm infant. *Pediatric Research*, 70(5), pp.507–512.

Thomas, E.L. et al., 2012. The effect of preterm birth on adiposity and metabolic pathways and the implications for later life. *Clinical Lipidology*, 7(3), pp.275–288.

Tremaroli, V. & Bäckhed, F., 2012. Functional interactions between the gut microbiota and host metabolism. *Nature*, 489, pp.242–249.

Trindade, C.E.P. et al., 2011. Fructose in fetal cord blood and its relationship with maternal and 48-hour-newborn blood concentrations. *Early Human Development*, 87(3), pp.193–197.

Trump, S. et al., 2006. 1H-NMR metabolic profiling of human neonatal urine. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 19(6), pp.305–312.

Tucker M, et al., 1991. Etiologies of preterm birth in an indigent population: is prevention a logical expectation? *Obstet Gynecol*. 77:343e7

Tucker, J. & McGuire, W., 2004. ABC of preterm birth Epidemiology of preterm birth. *BMJ*, 329, pp.675–678.

Turnbaugh, P.J. et al., 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 444, pp.1027–1031.

Turroni, F. et al., 2012. Diversity of bifidobacteria within the infant gut microbiota. *PLoS ONE*, 7(5), pp.20–24.

Uthaya, S. et al., 2005. Altered adiposity after extremely preterm birth. *Pediatric Research*, 57(2), pp.211–215.

van der Schoor, S.R. et al., 2004. Lysine kinetics in preterm infants: the importance of enteral feeding. *Gut*, 53(1), pp.38–43.

Vaz, F.M. & Wanders, R.J.A., 2002. Carnitine biosynthesis in mammals. *Biochem. J.*, 361, pp.417–429.

Vockley J, Ensenauer R. 2006. Isovaleric acidemia: New aspects of genetic and phenotypic heterogeneity.*AmJ Med Genet Part C Semin Med Genet* 142. pp95–103.

Voreades, N., Kozil, A. & Weir, T.L., 2014. Diet and the development of the human intestinal microbiome *Frontiers in Microbiology*. 5, p.494

Walhoud, K.B. et al., 2012. Long-term influence of normal variation in neonatal characteristics on human brain development. *Pnas*, 109(49), pp.20089–20094.

Waligora-Dupriet, A.J., et al., 2011. Diversity of gut Bifidobacterium species is not altered between allergic and non-allergic French infants. *Anaerobe*. 17(3): pp. 91-6.

Wang, L. et al., 2009. Is urinary indolyl-3-acryloylglycine a biomarker for autism with gastrointestinal symptoms? *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*, 14(8), pp.596–603.

Wang, L. et al., 2011. A review of candidate urinary biomarkers for autism spectrum disorder. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*, 16(7), pp.537–52.

Wang, Y. et al., 2009. 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *The ISME Journal*, 3(8), pp.944–954.

Wang, Z. et al., 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*, 472, pp.57–63.

Weir, T.L. et al., 2013. Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults. *PLoS ONE*, 8(8), p.e70803.

Wen, S.W. et al., 2014. Epidemiology of preterm birth and neonatal outcome. *Seminars in Fetal & Neonatal Medicine*, 9, pp. 429–435.

Weng, S.F. et al., 2012. Systematic review and meta-analyses of risk factors for childhood overweight identifiable during infancy. *Archives of disease in childhood*, 97(12), pp.1019–1026.

Wiley, W.C. & McLaren, I.H., 1955. Time-of-flight mass spectrometer with improved resolution. *Review of Scientific Instruments*, 26(12), pp.1150–1157.

Wilson, I.D. et al., 2005. High Resolution “Ultra Performance” Liquid Chromatography Coupled to oa-TOF Mass Spectrometry as a Tool for Differential Metabolic Pathway Profiling in Functional Genomic Studies research articles. *Journal of Proteome Research*, 4, pp.591–598.

Wojcik, Katherine Y. et al. 2009. Macronutrient Analysis of a Nationwide Sample of Donor Breast Milk. *J Aca of Nutrition and Dietetics*. 109 (1) pp.137 – 140

Wold, S., Esbensen, K. & Geladi, P. 1987. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*. 2, 37-52.

Wolfer et al., 2017. SANTA: Short Asynchronous Time-Series Analysis Using Smoothing Splines (In Press)

Woo, P.C.Y. et al., 2008. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *European Society of Clinical Infectious Diseases*, 14, pp.908–934.

Wu, J. et al., 2010. An optimised sample preparation method for NMR-based faecal metabonomic analysis. *The Analyst*, 135(5), pp.1023–1030.

Yamato, Y., et al., 2001. Fetal bile acid metabolism: analysis of urinary 3beta-monohydroxy-delta (5) bile acid in preterm infants. *Biol Neonate*. 80(1): pp. 19-25.

Yap, I.K.S. et al., 2010. Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. *Journal of Proteome Research*, 9(6), pp.2996–3004.

Younoszai, M.K. et al., 1969. Urinary hydroxyproline: creatinine ratio in normal term, pre-term, and growth-retarded infants. *Archives of disease in childhood*, 44(236), pp.517–520.

Zeisel, S.H. et al., 1989. Conversion of dietary choline to trimethylamine and dimethylamine in rats: dose-response relationship. *The Journal of nutrition*, 119(5), pp.800–804.

Zeisel, S.H., 2006. Requirements in Adults. *Annu Rev Nutr*, 26, pp.229–250.

Zeisel, S.H., Carolina, N. & Lien, E.L., 1996. Bioavailability of choline and choline esters from milk in rat pups. 2863(96), pp.457–464.

Zhang, C. et al., 2015. Dietary Modulation of Gut Microbiota Contributes to Alleviation of Both Genetic and Simple Obesity in Children. *EBIOM*, 2, pp.968–984.

Zhang, X. et al., 2015. Illumina MiSeq Sequencing Reveals Diverse Microbial Communities of Activated Sludge Systems Stimulated by Different Aromatics for Indigo Biosynthesis from Indole. *PLoS one*, 10(4), p.e0125732

SUPPLEMENTARY

Figure S1 – Dilution Series

Urine volume	Volume of water (ml)	Volume of phosphate buffer (ml)	Number of acquisitions performed
Volume 1: 540ml	0	60	32
Voume 2: 270ml	270	60	32
			128
Voume 3: 135ml	405	60	32
			512
Voume 4: 68ml	472	60	32
			2048

Figure S2 – Falcon centrifuge apparatus for urine contaminant investigation

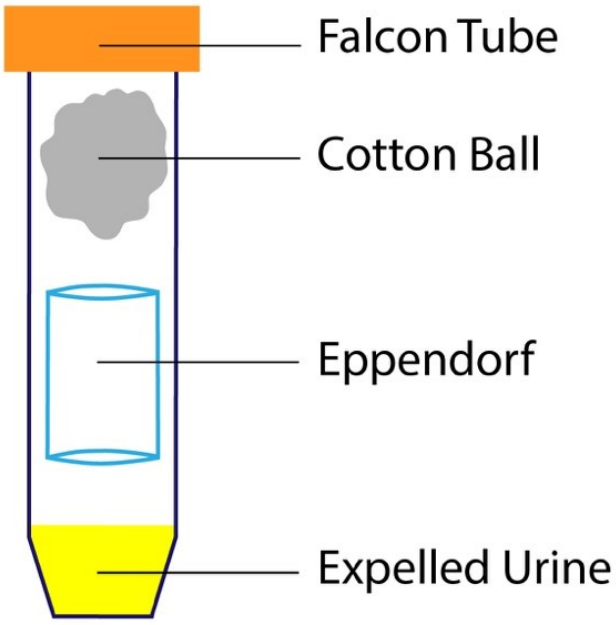


Figure S3 – Integral of TSP peak in increasing scans NMR analysis

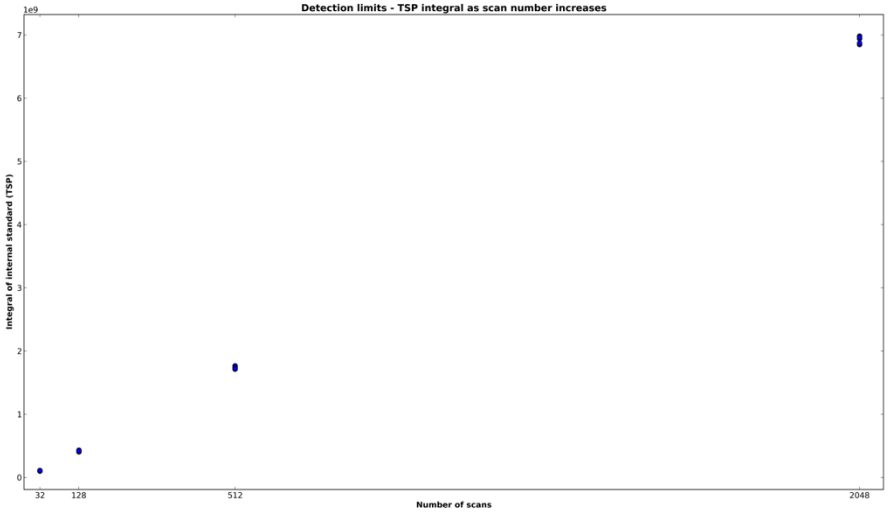


Figure S4 – PCA depicting faecal water acquired with different number of scans

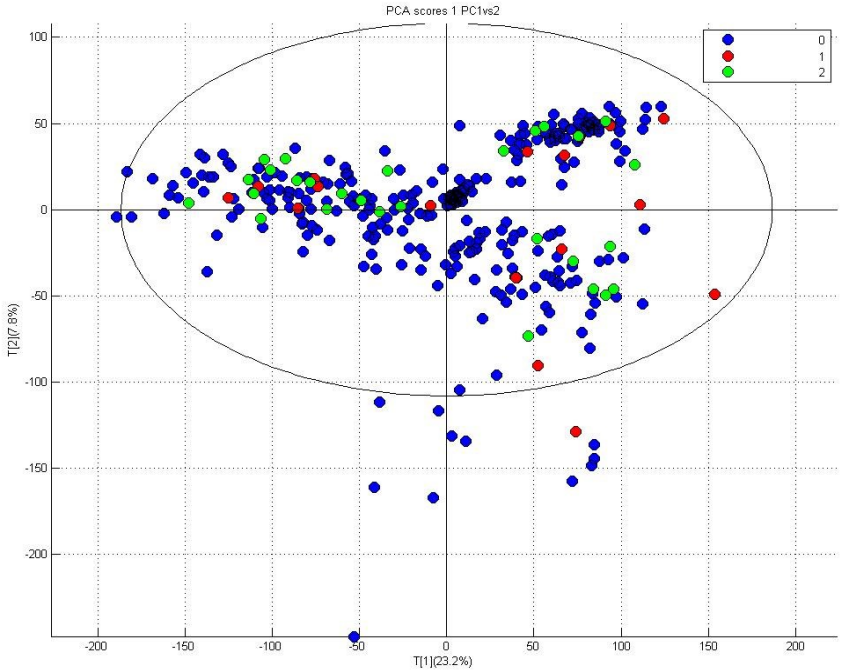


Figure S5 – PCA urinary NMR spectral data – Term at birth vs. Preterm at term equivalent age

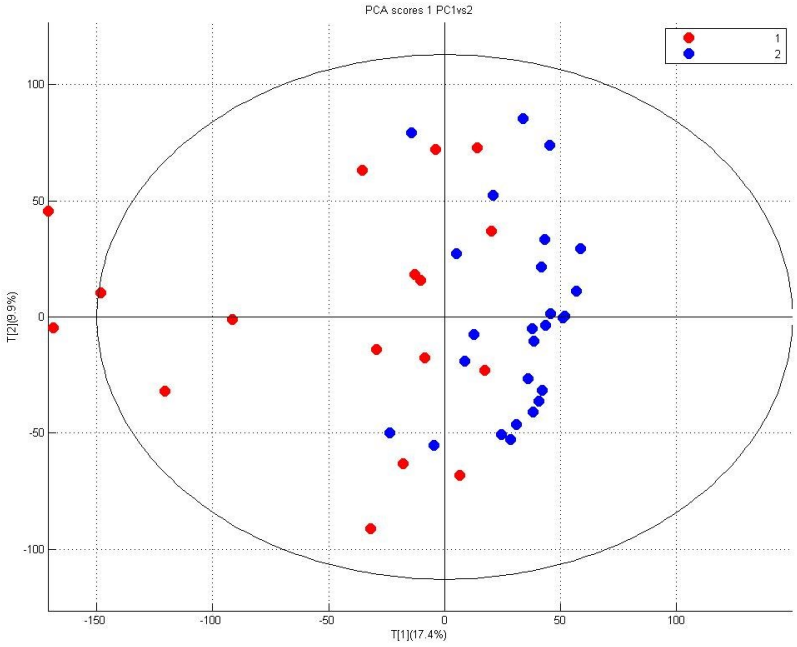
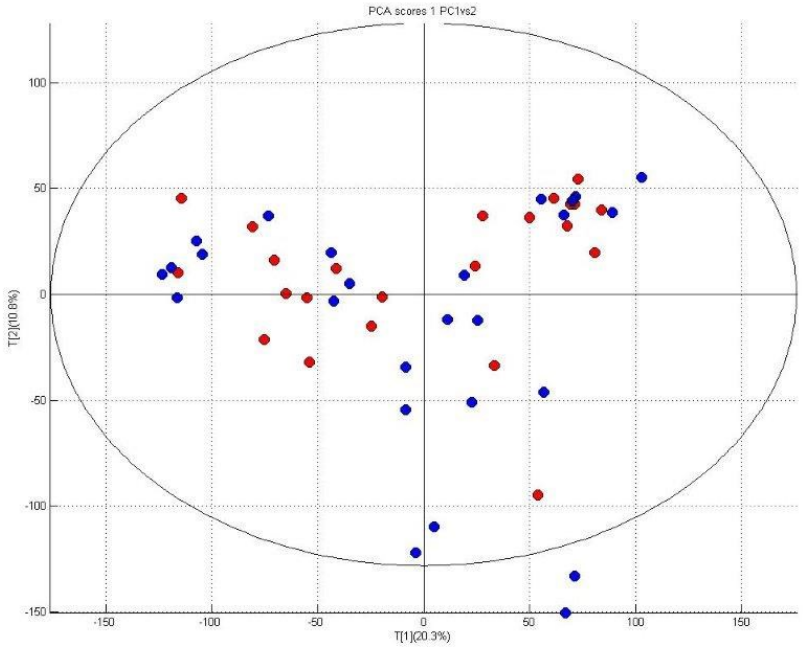


Figure S6 – PCA faecal NMR spectral data – Term vs. Preterm at 3 months postpartum



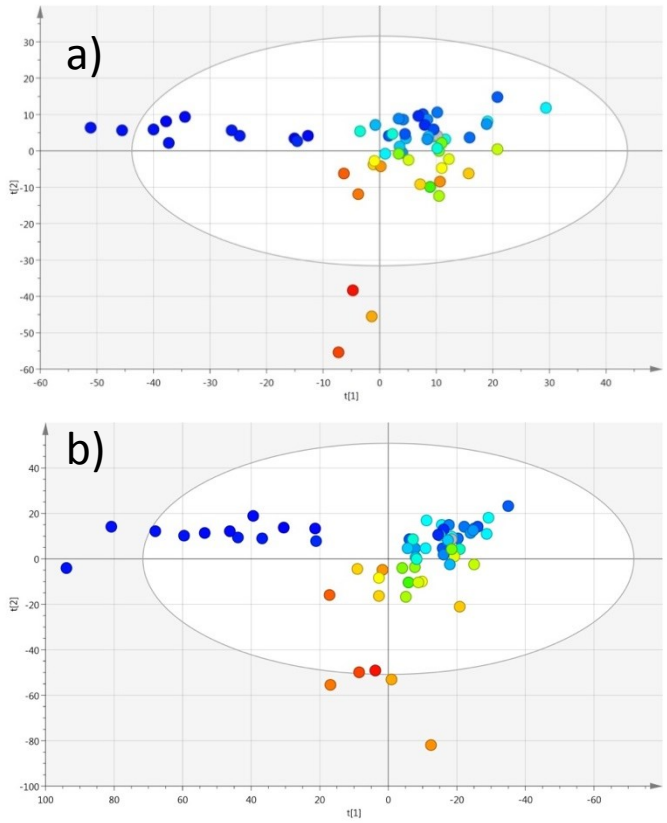


Figure S7 – HILIC-MS – Term vs Preterm coloured by increases postpartum age – blue through to red a) negative mode b) positive mode

Figure S8 – NoMic multivariate statistic values.

Model name	Model type	Time points	No. of samples	R2X (cum)	RYX (cum)	Q2 (cum)
Time point	PCA	ALL	1802	0.34		0.33
	OPLS	All	1802	0.32	0.78	0.73
Birth GA	OPLS	All	1802	0.34	0.44	0.22
		4	287	0.29	0.51	0.28
		10	288	0.33	0.84	0.59
		30	270	0.33	0.8	0.45
		120	234	0.27	0.57	0.17
		365	398	0.08	0.37	0.0044
		730	298	0.13	0.31	-0.12
Birth weight	OPLS	All	1802	0.32	0.34	0.13
		4	287	0.35	0.75	0.29
		10	288	0.33	0.79	0.44
		30	270	0.31	0.68	0.31
		120	234	0.24	0.35	0.13
		365	398	0.16	0.25	-0.07
		730	298	0.15	0.29	-0.17
Breastfeeding - exclusive	OPLS	All	1310	0.29	0.14	0.06
		4	188	0.23	0.37	0.02
		10	200	0.22	0.42	0.2
		30	187	0.24	0.44	0.23
		120	167	0.26	0.51	0.34
		365	314	0.17	0.23	0.091
		730	242	0.14	0.31	-0.17
Mode of Delivery	OPLS-DA	All	1784	0.31	0.15	0.05
		4	283	0.14	0.36	0.16
		10	284	0.25	0.49	0.16
		30	268	0.18	0.32	0.07
		120	231	0.22	0.32	-0.029
		365	393	0.081	0.35	-0.19
		730	299	0.14	0.38	-0.25
Term v Preterm	OPLS-DA	All	1801	0.27	0.15	0.09
		4	287	0.21	0.39	0.11
		10	288	0.2	0.49	0.31
		30	270	0.23	0.42	0.25
		120	234	0.26	0.56	0.19
		365	398	0.129	0.29	-0.006
		730	298	0.15	0.29	0.19

Appendix 1

Development of a Pipeline for Exploratory Metabolic Profiling of Infant Urine

Frances Jackson, Nancy Georgakopoulou, Manuja Kaluarachchi, Michael Kyriakides, Nicholas Andreas, Natalia Przysiezna, Matthew J. Hyde, Neena Modi, Jeremy K. Nicholson, Anisha Wijeyesekera, and Elaine Holmes

Journal of Proteome Research, 2016, *15* (9), pp 3432–3440

DOI: 10.1021/acs.jproteome.6b00234

Publication Date: August 1, 2016