Placental infection and maternal immunity in the Baby Bio Bank

Lydia Jane Leon

UCL

Thesis submitted for the degree of
Doctor of Philosophy

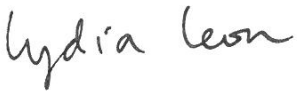
Genetics and Genomic Medicine
UCL Great Ormond Street Institute of Child Health

February 2017

Declaration

I, Lydia Jane Leon confirm that the work presented in this thesis is my own. Recruiters, employed as part of the Baby Bio Bank project, carried out participant recruitment, as well as clinical data and sample collection. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:



Lydia J Leon

Abstract

The Baby Bio Bank (BBB) is a UK based pregnancy complications cohort, with clinical data and biological samples from over 2500 mother, father, and infant trios. Within this large cohort, a significant number of infants were born preterm (<37 weeks gestation). The experimental work outlined in this thesis focused on these participants, alongside a selection of full term pregnancies for comparison.

All investigations were based on the hypothesis that many preterm births, particularly those precipitated by spontaneous events, have an underlying infectious aetiology. Total bacterial load in samples of placental DNA from 225 term and 141 preterm pregnancies was quantified, using quantitative PCR. An increase in total bacterial load was observed in placenta from spontaneous preterm births, compared to all other outcome groups.

A subsection of these samples were then taken forward for next generation sequencing of the 16S gene. We hypothesised that bacterial DNA found in the placenta would differ qualitatively, as well as quantitatively, between preterm and term pregnancies. An enrichment of genera previously associated with adverse outcomes, such as *Mycoplasma* spp., was observed in spontaneous preterm placenta. Other, less established oral species were also associated with this outcome, such as *Capnocytophaga* spp. Evidence for the existence of a normal, healthy placental microbiome was less clear, and these analyses were complicated by a significant impact of delivery contaminants on sequencing results.

The concentrations of 27 cytokines were also quantified in 149 maternal sera samples. These samples were from pregnancies that resulted in either term or spontaneous preterm births, and were taken within six weeks of delivery. Associations between cytokine concentrations and outcomes of interest were investigated. An upregulation of CC-type chemokines, as well as a number of other cytokines, was observed in spontaneous preterm pregnancies, implicating inflammatory events in the aetiology of spontaneous preterm birth.

Acknowledgements

I feel truly fortunate to have been under the wings of Professors Gudrun Moore and Philip Stanier, not just during this PhD, but throughout my life as a molecular biologist. I could not have asked for more supportive, kind, generous, and stimulating supervisors to help me become the scientist I am beginning to feel like today. Thank you both so very much.

Thank you to all my colleagues and friends in the lab for the advice, friendship, and distractions when I needed them most. Special thanks to Miho, who first showed me how to hold a pipette many moons ago, and has been the best desk neighbour and discussion buddy I could have asked for. Thank you to Sayeda, Pambos, Anna, Lara, Melissa, Veronica, Will, Dale, Kevin, Jess, Emma, Susanne, and Rimante for times shared in the lab, at lunch, and in the Lamb. I would also like to extend particular thanks to Nita Solanky, and the BBB recruiters. Thank you also to Chris Thalasselis, Lorraine Daly, Hemlata Varsani, and Mary Fewtrell.

I was fortunate enough to travel to the United States as part of my PhD, I am grateful to Professor Gregory Buck and Dr Jennifer Fettweis for this wonderful opportunity. Thanks too to Anita Marinelli, Ana Lara, Myrna Serrano, Vladimir Lee, and, my pal, Molly Johnson.

Discussions with Nigel Klein helped get this project off the ground, and I cannot extend enough thanks to Dr Ronan Doyle for all his help and generosity over the past few years. Thanks also to Kathryn Harris and Liam Shaw for helpful discussions. Thank you to Professor Taane Clark for his invaluable assistance on a draft, and for putting me in touch with Ernest Diez Benavente – a brilliant scientist and the most generous computational whizz I have come across.

To my wonderful friends, you are the best and you will never know how much your support has helped me get this job done. To Queensbridge, Coningsby, Henny, and all the other houses that have felt like home during this time, thank you for eating pasta, and talking about things other than DNA with me. To my sister in science, Georgia Pitts: let's go have some fun now. Thank you also to Dan, for so much, more than I think you know.

Mum, Dad, and Sam, thank you for all you have done, and continue to do, to support me and show me you believe in me. There aren't enough words to express my gratitude.

Lastly, my warmest thanks go to the BBB families, without whom none of this research would have been possible. This PhD was funded by the UK MRC, BRC, and GOSHCC.

Table of Contents

List of Figures

List of Tables

Abbreviations

°C	Degrees centigrade
AIC	Akaike Information Criterion
ANOVA	Analysis of variance
BLAST	Basic local alignment search tool
BMI	Body mass index
BV	Bacterial vaginosis
cm^3	Centimetres cubed
CS	Caesarean section
CT	qPCR cycle threshold
CVS	Chorionic villus sampling
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
ELISA	Enzyme linked immunosorbent assay
E	Scientific exponent
FDR	False discovery rate
fFN	Fetal fibronectin
FGR	Fetal growth restriction
g	Gravitational force
gm	Grams
GA	Gestational age
HCA	Histologic chorioamnionitis
HMP	Human microbiome project
Hz	Hertz
KO	Knock-out
k/mm^2	Thousand clusters per millimetre squared of Illumina flow cell
LOD	Limit of detection
LPS	Lipopolysaccharide
mg	Milligram
ml	Millilitre
NA	Not available
NaOH	Sodium hydroxide
NGS	Next generation sequencing
NICE	UK National Institute for Health and Clinical Excellence
NIH	USA National Institutes of Health
nM	Nanomolar
nsPTB	Non-spontaneous/iatrogenic preterm birth
OR	Odds ratio
OTU	Operational taxonomic unit
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Preeclampsia
pM	Picomolar
PTB	Preterm Birth
PPROM	Preterm premature rupture of membranes
QC	Quality control
qPCR	Quantitative polymerase chain reaction
RDP	Ribosomal Database Project
RM	Recurrent miscarriage

RPM	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
SA-PE	Streptavidin-phycoerythrin
SD	Standard deviation
SNP	Single nucleotide polymorphism
sPTB	Spontaneous preterm birth
sPTL	Spontaneous preterm labour
SROM	Spontaneous rupture of membranes
spp.	Species
USA	United States of America
VST	Variance stabilising transformation
WGS	Whole genome shotgun
µl	Microliter
X^2	Chi-squared
β	Beta, regression coefficient

Chapter 1: General Introduction

This thesis began as an exploration of a new clinical and biological resource for research into pregnancy complications within the UK: the Baby Bio Bank (BBB) (Abu-Amero et al., 2014, Leon et al., 2016). The BBB offers opportunities to explore four of the most common complications: preeclampsia (PE), preterm birth (PTB), fetal growth restriction (FGR), and recurrent miscarriage (RM), from many different molecular and clinical perspectives. After consideration of the practical possibilities of this resource and gaps in knowledge of pregnancy complications, I chose to focus my investigations on the molecular and microbiological aetiology of PTB within the BBB population.

As a student with a background in molecular genetics and fetal growth, and a personal interest in infectious disease, a chance to explore the nascent field of the maternal microbiome during pregnancy was both exciting and novel. Specifically, I chose to study whether the so-called 'placental microbiome', which can be defined as all of the microorganisms inhabiting this particular tissue, is altered in pregnancies that terminate in a PTB. In addition, were there identifiable signals of maternal immune disruption in PTBs from the BBB when compared to healthy, term births?

Clinical and molecular research into PTB has a long and broad history. The following general introduction will outline some of the key epidemiological, clinical, molecular, and technological developments that have contributed to our thorough, yet still incomplete, picture of the PTB phenotype. It will also touch on the tools currently available to clinicians for its prediction and prevention. On account of its known importance to PTB risk, and moreover the focus of this PhD, this introduction will pay particular attention to the role of intra-uterine infection in PTB. A more detailed exploration of specific developments and challenges concerning the maternal microbiome, and immunology and parturition, will be provided within the relevant individual chapters. A comprehensive review of all the multiple causes and consequences of PTB, is beyond the scope of this thesis. Furthermore, several excellent reviews already exist in the literature (Goldenberg et al., 2008, Blencowe et al., 2013, Romero et al., 2006, Iams et al., 2008).

1.1 Preterm birth (PTB)

PTB, defined as any delivery before 37 completed weeks of gestation, is the leading cause of neonatal morbidity and mortality worldwide, and inflicts substantial physical, psychological and economic costs upon society. It is best conceptualised as a syndrome rather than one homogenous disease phenotype. The clinical manifestations of PTB are varied, and much of its underlying aetiology remains incompletely understood (Romero et al., 2014a).

PTB can be split into two broad categories: spontaneous and indicated/non-spontaneous. Spontaneous PTB (sPTB) includes two further sub-types: spontaneous preterm labour (sPTL) with intact membranes and preterm premature rupture of membranes (PPROM) resulting in preterm delivery, either vaginal or by caesarean section (CS). In either case, the initiation of labour or membrane rupture is a spontaneous event, elicited by maternal and/or fetal biochemical changes. By contrast, indicated/non-spontaneous PTB (nsPTB) is the result of a clinical intervention on account of fetal distress or maternal conditions, such as preeclampsia. Such cases culminate in delivery by CS or, rarely, preterm induction of labour. sPTL accounts for around 40-45% of all PTBs, PPROM for 30-35%, with nsPTBs making up 25-30% of the burden (Goldenberg et al., 2008).

PTB can be further divided according to gestational age (GA) at the time of occurrence. This division is associated with changing incidence, as well as mortality and morbidity rates. 'Very early PTB' is often defined as any birth before 33 weeks gestation and includes a high proportion of infants of very low birthweight (<1500 grams (gm)). This group represents 1-2% of all births. A higher number of these early births is spontaneous rather than indicated. The majority of serious illness and mortality is concentrated amongst these early deliveries (Goldenberg et al., 2000). In England and Wales, where some of the most advanced neonatal facilities are available for preterm infants, around 85% of infants born under 24 weeks die within the first year of life (Office for National Statistics, 2015). This percentage rapidly declines to 24% under 26 weeks and under 10% by 28 weeks gestation. By 34 weeks over 99% of infants survive. This significant relationship between GA at birth and survival highlights the importance of research identifying causal factors that precipitate the very earliest births.

1.1.1 Epidemiology of preterm birth

PTB affects between 5-18% of births worldwide. In the majority of countries with reliable data, PTB rates rose globally in the decades following the 1980s (Blencowe et al., 2013). This trend was especially marked in the United States of America (USA) where there was a 30% rise in the PTB rate between 1981 and 2006, reaching a peak of 12.8% (Martin and Osterman, 2013). Interestingly, although the USA remains one of the countries with the highest PTB rates, the latest figures indicate that since this peak, the rate has been declining and in 2013 was 11.4% (Martin et al., 2014). In terms of absolute numbers, less economically developed countries bear by far the largest burden of PTB incidence, morbidity and mortality. It is reported that approximately 85% of PTBs occur in Africa and Asia (Beck et al., 2010). The PTB rate in England and Wales is at the lower end of the global scale at around 7%, and has been steady over the past decade (Office for National Statistics, 2014, Office for National Statistics, 2015).

In industrialised countries, PTB incidence is strongly associated with maternal ethnicity (Khalil et al., 2013). In the UK, mothers of Caribbean and African origin have the highest rates of PTB, followed by Asian women, whilst white women have the lowest rates (Moser et al., 2008). In the USA, black women have almost twice as many PTBs as white women (Muglia and Katz, 2010). The above associations remain when data on potential confounding variables, such as income and education are taken into account (Goldenberg et al., 1996). PTB is a complex phenotype with a wide range of underlying, and often interacting causes. Table 1.1 outlines a selection of additional environmental, maternal, and physiological factors thought to contribute to the risk of PTB.

Table 1.1 – Previously identified risk factors for PTB.

Risk factor for PTB	Example reference(s)
Smoking	Goldenberg et al. (2008)
Maternal BMI	Shaw et al. (2014), Torloni et al. (2009), Girsen et al. (2016)
Maternal ethnicity	Schaaf et al. (2013)
Maternal stress	Petraglia et al. (2010)
Intra-uterine infection	Goncalves et al. (2002)
Variation in vaginal flora	Leitich et al. (2003)
Pre-eclampsia	Meis et al. (1995)
Inter-pregnancy interval	Conde-Agudelo et al. (2006)
Pregnancy history	Boyd et al. (2009), Carr-Hill and Hall (1985)
Maternal age	Restrepo-Mendez et al. (2015)
Maternal diabetes	Kock et al. (2010)
Genetic variation	Plunkett and Muglia (2008)
Parity	Kozuki et al. (2013)

1.1.2 Impact of preterm birth

An estimated 2.761 million neonatal deaths occurred globally in 2013, comprising 44% of all deaths under five (Liu et al., 2015). PTB is the largest contributor to these figures, accounting for around 29% of the total (Liu et al., 2015, Lawn et al., 2010), although survival rates following a PTB show substantial inter-country variation (Blencowe et al., 2013). Unlike other contributors to neonatal mortality such as tetanus, limited, if any, progress has been made in reducing the global mortality burden of PTB (Lawn et al., 2010). Furthermore, it is customary to record deaths among preterm infants that ultimately die from infection as infectious related deaths, rather than preterm complications. Since many of these neonates will have succumbed to the infectious insult as a direct result of their prematurity, the total contribution of PTB to neonatal mortality is likely to be underestimated in these figures.

Preterm infants who survive beyond the neonatal period often face a lifetime of increased health risk and complex morbidities (Saigal and Doyle, 2008). Compared with children born at term, premature infants have higher rates of infection, neurodevelopmental disorders, sensory deficits, respiratory and gastro-intestinal problems, and cerebral palsy. Even those infants born at later, 'safer' GAs (32-36 weeks), who do not generally require neonatal intensive care, face higher short and long term morbidity compared to their term peers (Blencowe et al., 2013).

Beyond the emotional and physical costs to the infant and their family of being born too early, there are substantial economic costs associated with PTB related care. In a recent study, Mangham et al. (2009) estimated that the total cost of a year's worth of PTBs within England and Wales, from birth to adulthood (18 years), was £2.946 billion. One third of these costs were borne during the neonatal period. Unsurprisingly, the cost of treating a neonatal infant was found to be inversely proportional to GA at birth. The model estimated that delaying PTB by one week, across all gestational categories, could save around one billion pounds in healthcare expenditure. This finding highlights the value of treatments that delay the onset of labour and keep the fetus *in-utero* for as long as is safely possible.

1.1.3 Predicting preterm birth

Identifying women at risk of PTB enables the implementation of appropriate observational, surgical or pharmacological strategies to minimise risks to the mother

and baby. In general, these serve to either prevent or delay labour/delivery, or reduce the neonatal risks associated with prematurity, such as immature organ development. However, identifying which pregnancies will result in a preterm delivery, and may benefit from the available therapies, remains a significant challenge. This is the case even amongst women showing symptoms of preterm labour. For example, in a study following women from 20-34 weeks gestation, who presented with uterine contractions and minimal cervical dilation, only 7% delivered before 34 weeks and 24% before 37 weeks. Therefore, 69% of women who threatened labour before 34 weeks ultimately delivered at term (Guinn et al., 1997).

One of the strongest predictors of PTB, and one of the only pre-pregnancy indicators, is a history of previous PTB (Mercer et al., 1999, Carr-Hill and Hall, 1985, Iams et al., 1998). Mercer et al. (1999) found that if a woman had previously experienced a sPTB they were 2.5 times more likely to experience another in the current pregnancy, when compared to women with no prior sPTB history. The size of this effect was inversely proportional to the GA of the previous sPTB. Furthermore, the type of sPTB (sPTL or PPROM) in the previous pregnancy correlated with outcome (sPTL or PPROM) in the current pregnancy. This relationship implied a preservation of pathological mechanisms between pregnancies, which were specific to sub-types of the syndrome.

Although difficult to assess accurately, estimates of heritability of the PTB phenotype range from 25% to 40% (Allen and Founds, 2013). In a large Norwegian cohort, mothers born preterm had a 55% increased chance of having a preterm delivery themselves. This risk increased when the mothers had been born at the more extreme end of prematurity (Wilcox et al., 2008). Interestingly, little or no evidence was observed in this study for the effect that preterm fathers had on the GA at birth of their children, implicating the maternal genotype rather than the paternal genotype in these repetitive events. Increased risks associated with individual pregnancy history, in addition to inter-generational effects, indicate an important role for specific genetic factors in PTB risk. A number of genes and single nucleotide polymorphisms (SNPs) have indeed been associated with PTB risk, many of which are associated with immunological pathways (Crider et al., 2005, Allen and Founds, 2013). It is also possible that the repetition of preterm labour within individuals could reflect the persistence of non-heritable risk factors such as chronic, potentially subclinical, infection. This latter hypothesis is rarely explored in the literature.

Among the most powerful predictive tests currently available to clinicians during pregnancy is the fetal fibronectin (fFN) test, which was first described in 1991 (Lockwood et al., 1991). The test measures the presence of the fFN protein in cervico-vaginal secretions. fFN is a placental glycoprotein present in amniotic fluid, fetal connective tissues, and placenta, that functions as an adhesion molecule, 'gluing' the placental tissue and membranes to the maternal decidua (Matsuura et al., 1988). Goldenberg et al. (1998) showed that a positive fFN test was the single strongest available indicator for PTB risk in *asymptomatic* women, and that this value was increased in earlier gestations. Following a positive fFN test, a relative risk of 14.1 (9.3-21.4 95% confidence interval (CI)) was calculated for preterm births before 32 weeks gestation and 6.7 (4.9-9.2 95% CI) for delivery before 35 weeks (Goldenberg et al., 1998). The study also reported that among asymptomatic women, risk of delivery before 37 weeks was highest amongst women with a history of PTB, a positive fFN result, *and* a shortened cervix.

Trans-vaginal ultrasound measurement of cervical length has been used as a method to predict PTB since the 1980s (Grimes-Dennis and Berghella, 2007). Many studies have shown that women with a shortened cervix, often defined as <25mm, in the mid-trimester of a singleton pregnancy, who may or may not have been threatening labour, are at a higher risk of delivering preterm (Grimes-Dennis and Berghella, 2007, Hassan et al., 2006, Botsis et al., 2005, Iams et al., 1996). However, the reported predictive value of this test varies considerably between studies and populations. Evidence suggests that cervical screening has greatest specificity in high-risk women, e.g. those with a history of a PTB (Grimes-Dennis and Berghella, 2007). It is generally hypothesised that these associations are the result of structural inadequacy of the shortened cervix. However, a reduced distance between the uterus and the bacterial rich vagina could also increase the chances of ascending infections reaching the intra-uterine cavity. The relative contribution of either hypothesis is under-explored. Correlations between intra-uterine infection and/or inflammation and shortened cervices have been reported. However, the direction of causality here remains unclear (Guzman et al., 1999, Goldenberg et al., 2000).

1.1.4 Preventing preterm birth

Cervical cerclage and, less commonly, the administration of progesterone can be used to delay or prevent premature labour in at-risk groups. Corticosteroids and tocolytics

can be offered to women threatening preterm delivery to delay labour and reduce prematurity associated risks to the fetus, such as respiratory distress and necrotizing enterocolitis. Despite their availability, evidence for the efficacy of these strategies is mixed, and clear consensus regarding the appropriate indications for their use is lacking.

Cervical cerclage, in which a stitch is placed around the cervix to keep it closed, was a technique first performed in 1902 (RCOG, 2011). It is now a commonly used prophylactic intervention in modern obstetrics for women with shortened and/or prematurely dilated cervices. A recent Cochrane review showed an appreciable reduction in PTB rates, but no impact on associated neonatal morbidities, when cerclage was performed (Alfirevic et al., 2012). Recent National Institute for Health and Clinical Excellence (NICE) accredited guidelines state that, dependent upon clinical examination, women with a history of multiple PTBs, or those with an ultrasound identified shortened cervix and PTB history, should be considered for this therapy (RCOG, 2011).

Progesterone is a critical hormone throughout pregnancy and its functional or systemic withdrawal towards the end of gestation has been hypothesised to play a central role in the initiation of labour in mammals (Golightly et al., 2011). These observations have highlighted progesterone as a potential treatment for women at risk of PTB. Although many trials showed positive effects, the overall evidence remained mixed (Meis and Aleman, 2004). The publication in 2003 of positive results from two large randomised controlled trials of its mid-gestational use in high-risk women, served to renew interest in its use (Meis et al., 2003, da Fonseca et al., 2003). Meis et al (2003) found that in women with a prior PTB, administration of the progestational compound 17α-hydroxyprogesterone caproate reduced the incidence of PTB by up to one third. A recent Cochrane review has provided further evidence for its reductive effect on PTB rates in at risk (short cervix or PTB history), singleton pregnancies. However, consensus over its use, dosage, and long-term effects, is yet to be reached (Dodd et al., 2013). Trials of progesterone as a tocolytic drug to delay labour have shown little, if any, efficacy (Meis and Aleman, 2004).

Given the strong evidence for a link between PTB and a variety of infections, there has been substantial interest in the use of antibiotics to prevent or delay early delivery. However, the use of antibiotic therapy in women showing signs of infection during

gestation, or threatening preterm delivery, has shown mixed success (Hutzal et al., 2008, Kenyon et al., 2013). A large body of clinical research has investigated their impact on pregnancy and neonatal outcomes, using both broad range and targeted drugs at a variety of GAs. Decisions regarding which drugs to use, in which situations, remain complex and controversial (Subramaniam et al., 2012).

Despite decades of research, and the widespread use of clinical strategies to prevent or delay PTB, there is little evidence that its incidence has shown any appreciable decline. Indeed, in many countries, rates appear to be rising (Blencowe et al., 2013). Furthermore, whilst neonatal survival, particularly in areas with access to modern medical techniques has improved considerably, reductions in associated morbidities have not been as significant (Stoll et al., 2015). Although this reduction in PTB-associated mortality is, of course, to be celebrated, these concomitant morbidity trends imply an absolute increase in PTB associated complications. New clinical and ethical challenges are arising in this context as the so-called 'limit of viability' continues to decline (Seri and Evans, 2008). Taken together, these data continue to highlight the importance of identifying clinical strategies that will improve neonatal outcomes by extending gestation for as long as is safely possible.

If prediction and prevention of PTB is to improve, we need a better understanding of the underlying pathology of this syndrome, including the molecular mechanisms that result in a woman embarking on the path to parturition too early. Some of the most promising mechanistic research of recent decades, and the focus of the experimental work for this thesis, has involved associations between PTB, and maternal infection and inflammation in pregnancy.

1.1.5 Pregnancy, infection and PTB

The danger of acute and chronic maternal infection during pregnancy is well known. Public health initiatives such as the provision of the flu vaccine to pregnant mothers, or development of specific food hygiene guidelines for pregnancy to avoid maternal ingestion of hazardous bacteria such as *Listeria monocytogenes* or the *Toxoplasma gondii* parasite, serve to mitigate the risks associated with certain infections in pregnancy (Plotkin, 2006, Steinhoff et al., 2012, Delorme-Axford et al., 2014). Maternal HIV and Hepatitis B infections have been associated with increased odds of PTB (Slyker et al., 2014, Reddick et al., 2011). Malaria has also been associated with shorter

gestations, particularly in the latter part of pregnancy (De Beaudrap et al., 2013, Uneke, 2007). A number of emerging infections, most recently in the case of the Zika virus, have continued to keep infection and pregnancy central to the global public health agenda, with particular emphasis on associated developmental pathologies such as microcephaly (Johansson et al., 2016). The following section will focus on bacterial infection in pregnancy, specifically of the female reproductive and intra-uterine tissues, and their associations with PTB. Whilst viral, parasitic, and fungal infections also contribute to fetal morbidity and pregnancy complications, the varied clinical and molecular pathology of all these diverse groups of organisms is beyond the scope of this thesis.

Identification of bacteria from within the normally 'sterile' amniotic cavity were reported as early as the 1920s (Harris, 1927). However, molecular evidence causally linking such colonisation to PTB was not published until the 1970s (Bobitt and Ledger, 1977). In the decades since, functional and observational evidence has continued to grow in support of the hypothesis that bacterial infection during pregnancy, particularly of the maternal reproductive tissues, is a significant contributor to PTB incidence. It is estimated that somewhere between 25% and 40% of sPTB cases may be mediated by intrauterine infection (Goldenberg et al., 2008). This proportion increases steadily as GA at birth decreases. Infection may be a mediator in as many as 79% of births at 23 weeks gestation (Onderdonk et al., 2008a).

Evidence for the involvement of bacteria in preterm delivery comes from the identification of organisms within pregnancy related tissues using a variety of techniques including histopathology, immunohistochemistry, protein-assays, bacterial culture, and polymerase chain reaction (PCR) based methods. Such techniques have helped demonstrate an association between sPTB and the presence or profile of bacteria in amniotic fluid (Han et al., 2009, Hatanaka et al., 2014, Wang et al., 2013), placental parenchyma (Aagaard et al., 2014, Onderdonk et al., 2008b), fetal membranes (Fortner et al., 2014, Jones et al., 2009), the placental basal plate (Stout et al., 2013), cord blood (Wang et al., 2013), cervical fluid (Musilova et al., 2014), and maternal vaginal flora (de Andrade Ramos et al., 2014). The association between bacterial vaginosis (BV), in which a disruption in maternal vaginal flora leads to a loss of *Lactobacilli* species (spp.) and overgrowth of facultative anaerobe species, and PTB has been acknowledged for some time (Hillier et al., 1995). A meta-analysis of 18 studies including over 20,000 women, estimated that women with BV in early pregnancy had a

2 fold higher odds of PTB than asymptomatic women (Leitich et al., 2003). Although these observations provide support for the contribution of the maternal reproductive microbiome to PTB risk, BV is a fairly common pathology and its predictive power is limited, most women with BV will deliver at term (Romero et al., 2014a).

Typically, organisms found in the intra-uterine tissues are hypothesised to originate in the vagina, ascend the genital tract, breaking the cervical barrier and colonising all or some of the amniochorionic membranes, the amniotic fluid, the parenchyma and villous tissue below the placental membranes, the umbilical cord and, rarely, the fetus itself (Figure 1.1). However, some recent evidence has pointed to alternative routes of infection, such as via haematogenous transport of bacterial species originating from non-reproductive tissues such as the oral cavity (Aagaard et al., 2014, Han and Wang, 2013, Offenbacher et al., 2006). Infection of reproductive tissues in pregnancy may also be the result of invasive clinical procedures such as amniocentesis, chorionic villus sampling (CVS), and perhaps even cervical cerclage. Recent experimental work has also raised the possibility that in some (or all) women, bacteria may even be present in non-pregnant uterine tissues prior to implantation (Verstraelen et al., 2016).

All women harbour significant numbers of microorganisms within their reproductive tracts, with reproducible shifts in community composition noted before, during, and after pregnancy (MacIntyre et al., 2015). In the majority of cases these organisms will never reach the amniotic cavity, or its surrounding tissues. Even fewer will elicit a pathogenic response to their presence that may result in PTB. Many adaptations exist that serve to protect women from ascending infection. The mucus plug that accumulates at the cervix during pregnancy acts as a key physical and functional barrier to ascending infection. The release of anti-microbial, and pro-inflammatory peptides by vaginal epithelial cells additionally helps to create an environment hostile to overgrowth of non-commensal bacterial species (Witkin, 2014). However, in the rare cases in which these defences are compromised, or indeed bacteria and/or their products reach the uterus via the maternal bloodstream, a number of mechanisms have been postulated by which this bacterial infection may act to initiate or increase the risk of PTB. Most hypotheses about the mechanisms involved in infectious mediated PTB, regardless of where the organisms are thought to originate, propose a central role for inflammation-related pathways.

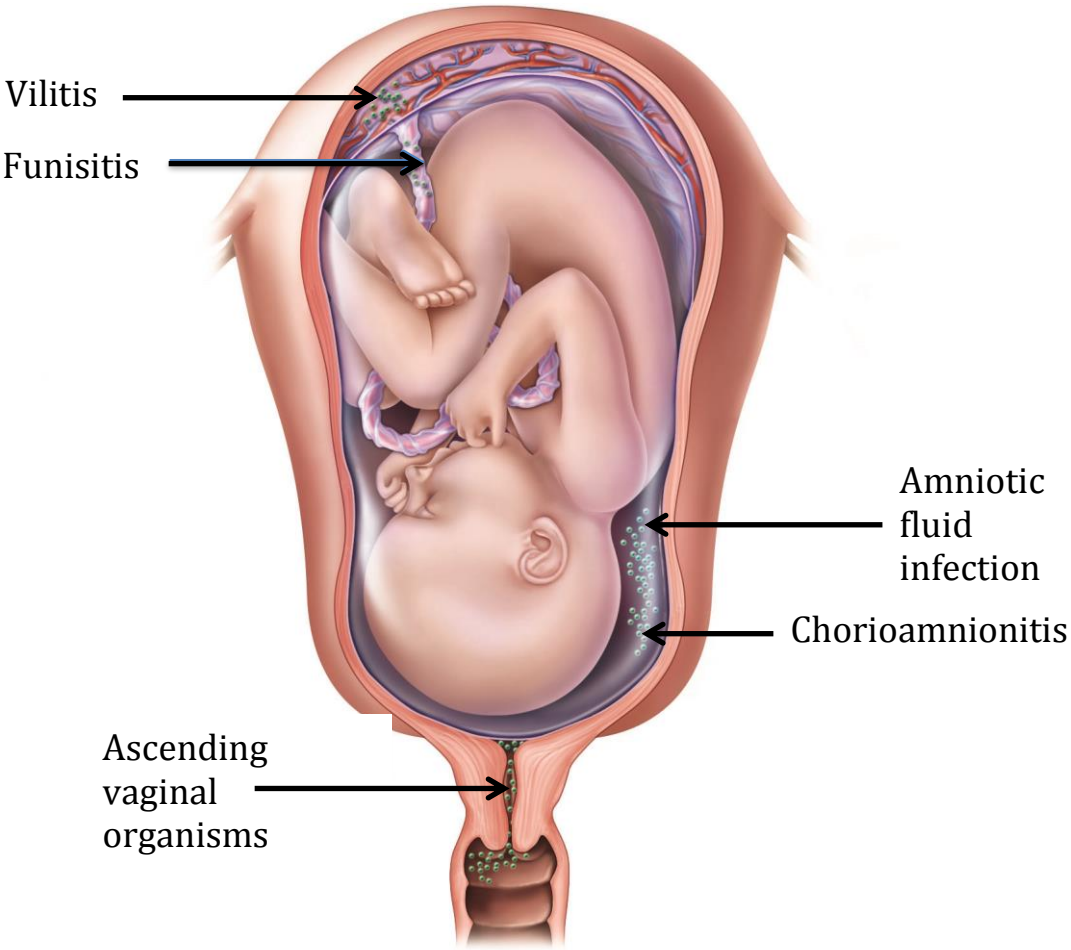


Figure 1.1 – Potential route of infection and reported sites of intra-amniotic infection. *Image used with permission from Suzanne Ghuzzi-Silva, MPS North America (sghuzzi@gmail.com).*

1.1.6 Inflammation and chorioamnionitis

The role of inflammation in infectious mediated PTB is strongly implied by the widely reported association between sPTB and intrauterine inflammation, most commonly in the form of clinical or subclinical (histologic) chorioamnionitis (Guzick and Winn, 1985, Mueller-Heubach et al., 1990, Kim et al., 2015). This condition is usually defined as inflammation of the fetal membranes (the outer chorion and inner amnion), which is mainly in the form of maternal neutrophilic infiltration (Figure 1.2). The majority of cases are histologic chorioamnionitis (HCA) with no observable clinical phenotype. Some are also associated with uterine tenderness, leucocytosis, fever and tachycardia and are known as clinical chorioamnionitis, but this proportion may be as low as 8% of the total (Guzick and Winn, 1985). Less commonly, inflammation may also affect the villous tree, known as villitis. Funisitis, in which inflammation of the umbilical cord is

observed, tends to be the result of fetal, rather than maternal, inflammatory responses (Kim et al., 2015).

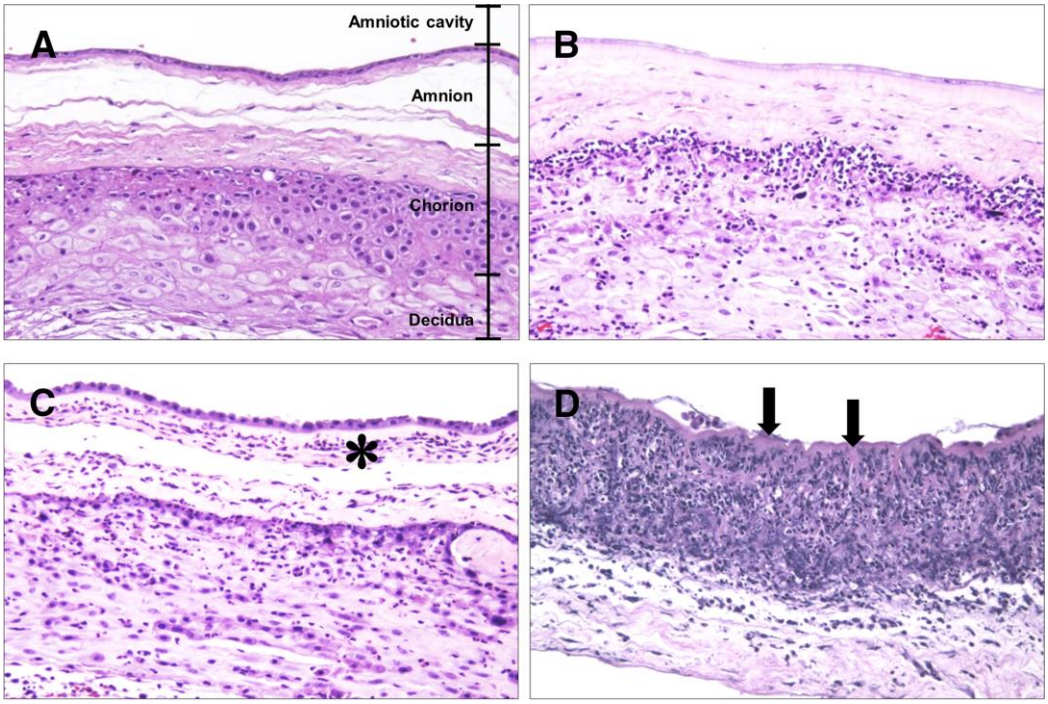


Figure 1.2 - Acute chorioamnionitis of the extraplacental chorioamniotic membranes: A, Normal chorioamniotic membranes shows the absence of neutrophils. **B,** Acute chorionitis is stage 1 acute inflammation of the chorioamniotic membranes, in which neutrophilic infiltration is limited to the chorion. **C,** Acute chorioamnionitis is stage 2 acute inflammation of the chorioamniotic membranes; neutrophilic migration into the amniotic connective tissue is shown (*asterisk*). **D,** Necrotizing chorioamnionitis is stage 3 acute inflammation of the chorioamniotic membranes, whose characteristic is the amnion epithelial *necrosis (*arrows). *Image and legend reprinted from (Kim et al., 2015) with permission from Elsevier.*

Incidence of chorioamnionitis is negatively correlated with GA at birth. Some studies report HCA incidence of over 90% in the earliest (21-24 weeks) deliveries (Kim et al., 2015). Chorioamnionitis is often, but not exclusively, accompanied by positive results from bacterial cultures or PCR tests (Hillier et al., 1988, Redline, 2004, Galinsky et al., 2013). The observed associations between chorioamnionitis, bacterial presence, and GA at birth, provide a plausible mechanistic link between intra-uterine bacterial infection and PTB. These associations implicate the maternal, or fetal, inflammatory responses to infection as possible triggers for labour and/or membrane rupture.

Intrauterine infection is rarely accompanied by clinical symptoms; at least until membranes break or labour begins. This makes accurate estimation of the burden of PTB attributable to infection difficult. Furthermore, until recently, diagnosis of intra-

amniotic infection relied either on histological evidence of chorioamnionitis or microbiological cultures. The reliance on these two strategies may have led to an underestimation of the infectious mediated PTB burden. Firstly, chorioamnionitis does not occur in all cases of infectious mediated PTB (Goncalves et al., 2002). Secondly, culturing techniques are only able to identify the presence of a limited proportion of microorganisms. This means that certain intra-amniotic infections, particularly novel ones, may be overlooked.

Evidence for causality in the observed association between bacterial infection, inflammation, and PTB has come in large part from animal studies (Elovitz and Mrinalini, 2004). Many experimental models have shown that exposure to both live bacterial cells, as well as bacterial products, such as the endotoxin lipopolysaccharide (LPS), can induce host immune responses. These immune responses serve to trigger preterm labour, in species as diverse as rhesus monkeys (Gravett et al., 1994), sheep (Schlafer et al., 1994) and mice (Mussalli et al., 1999, Elovitz et al., 2003). Figure 1.3 shows the likely physiological sequence of events linking bacterial infection to PTB in humans. The immunology of normal and pathological labour will be considered in more detail in Chapter 7.

The precise mechanism by which infection results in PTB will depend on many factors. The origin of bacteria, e.g. oral or vaginal, will implicate what species infect the intra-uterine space, and therefore what virulence factors elicit maternal or fetal responses (Fichorova et al., 2011, Menon et al., 2009). Where organisms originate will also impact which intra-uterine tissues become colonised in the first place. The maternal and fetal immune responses to the infectious insult are likely to vary on account of individual genetic factors, as well as variations in bacterial virulence factors. This variation may trigger different physiological pathways that result in labour, and may also impact neonatal outcomes differently. Additionally, at what stage of pregnancy the infection takes hold will be an important variable in the determination of both maternal and fetal responses to the insult, as well as maternal and fetal health.

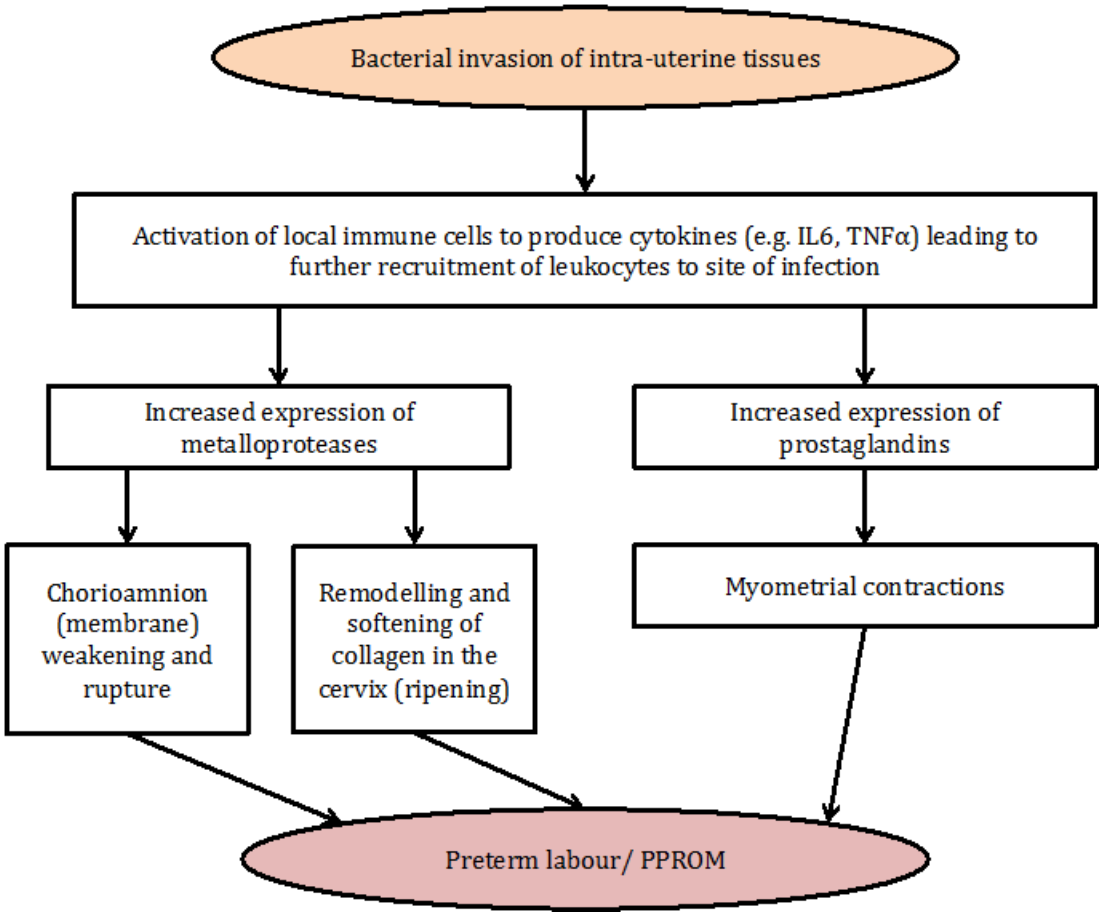


Figure 1.3 - Hypothesised physiology of infectious mediated PTB, adapted from Goldenberg et al. (2000).

1.1.7 Infection and the placenta

Throughout pregnancy, the placenta is the organ through which all necessary exchanges of gases, nutrients and waste products take place. In addition, the placenta serves as a functional and immunological barrier from the hematogenous spread of microorganisms from mother to fetus, as well as from maternal immune system attacks (Delorme-Axford et al., 2014, Nelissen et al., 2011). The placenta consists of both fetal and maternally derived tissues that fuse together in early pregnancy to facilitate diffusion of substances between the maternal and fetal circulatory systems. The organ develops into three major structures: the umbilical cord, the chorioamniotic membranes, and the placental disc itself, which consists of parenchyma tissue on the fetal side and villous tissue at the maternal edge. The placenta's functional capacity is a critical determinant of proper fetal growth and development.

Although evidence exists for the presence of bacteria in all intra-uterine tissues, the majority of studies have focused on identifying bacterial colonisation of the fetal membranes and amniotic fluid. The prevailing view has been that 'bacterial infection within the placenta (villitis) is rare' (Goldenberg et al., 2000). Nevertheless, evidence exists for the association between colonisation of placental tissue and adverse pregnancy outcomes (Onderdonk et al., 2008a, Onderdonk et al., 2008b, Prince et al., 2016, Rours et al., 2011). Estimating the incidence and precise nature of pathogenic placental infection is difficult. Histopathological investigations of placental tissue can necessarily only identify infection *after* the event. Furthermore, evidence suggests that bacteria may be present in the uterus without causing a clinically observable inflammatory response (Steel et al., 2005). The organ cannot be studied at any time other than after birth, other than in extenuating clinical cases, such as when CVS is offered to mothers at risk of chromosomal abnormalities. By the time a placenta is delivered, the infection may have been proliferating, shrinking, and interacting with the maternal and /or fetal immune systems for up to 9 months. At present, the vast majority of research documenting such infections can only be carried out retrospectively. Furthermore, tissue samples collected after birth can easily be contaminated with allochtonous organisms at delivery, precluding the identification of only those species actually present in the placental tissue *in-utero*.

Following the identification of bacteria in placental tissue from uncomplicated pregnancies (Stout et al., 2013) and the recent description of a so-called 'placental microbiome' (Aagaard et al., 2014), the dogma of placental sterility in normal pregnancy is being challenged. These observations have raised the possibility that the presence of bacteria in placental tissue may not always be damaging. Some have suggested it could represent an adaptive mechanism to prime the fetal immune system before birth (Satokari et al., 2009, Romano-Keeler and Weitkamp, 2014). The presence of bacteria in the meconium of newborn infants (Hansen et al., 2015) does indeed imply fetal exposure to certain organisms pre-delivery. However, the molecular detail of how placental colonisation could contribute to healthy fetal development remains sparse.

The above hypotheses have not gone unchallenged (Kliman, 2014). In the nascent world of microbiome research it is becoming increasingly clear that the distinction between noise, whether technological error or sampling contamination, and signal, is a challenge that has not always been appropriately addressed. This issue is a particular concern when investigating samples of low biomass such as the placenta. Indeed, a

recent paper has challenged the idea of a placental microbiome entirely, attributing all bacterial reads within their samples to background contamination (Lauder et al., 2016). In addition, it is impossible to distinguish competent, living bacterial cells, from dead cells or translocated, cell-free bacterial DNA, when using sequencing based techniques. These are critical and often poorly documented aspects of microbiome studies that will be considered within the context of my own research in the forthcoming chapters.

1.1.8 Characterising intra-uterine infection

An inherent challenge in PTB research is identifying pre-clinical signs of bacterial infection that will ultimately provoke a pathological pregnancy outcome. Procedures, such as amniocentesis, which give rare access to the intra-amniotic environment during pregnancy, are risky and unethical to conduct without good clinical reason. Post-natal investigations of fetal and maternal tissues are no longer relevant for clinical management to delay or prevent PTB, but have been useful for research. It is encouraging that new molecular techniques are helping to broaden our understanding of infectious mediated PTB. These are uncovering which specific strains of bacteria are implicated in these syndromes, and *where* such organisms tend to originate, thus giving clues to where eventual effective treatments could be targeted. However, almost all feasible study designs still rely on retrospectively collected samples to infer the existence of infection during pregnancy.

Traditionally, identification and characterisation of prokaryotic organisms have relied on the ability of the microbiologist to isolate and grow cells of interest under laboratory conditions. With the advent of molecular approaches in microbiological research, it soon became clear that culturing techniques had only been able to describe a fraction of global microbial diversity (Hugenholtz, 2002). This phenomenon is known as 'the great plate count anomaly' (Staley and Konopka, 1985). The use of PCR based molecular techniques in PTB research has confirmed such an underestimation of diversity. Studies have consistently shown that when the presence of bacteria in intra-uterine tissues is assessed using PCR and culture, culture-based methods are less sensitive and less able to identify mixed microbial infections (Han et al., 2006, Han et al., 2009, Harris and Hartley, 2003, Wang et al., 2013).

1.1.9 The 16S ribosomal RNA locus

Since Leuwenhoek's 17th century revolutionary developments of microscopy, identification and classification of microorganisms have largely been processes based on morphological and metabolic characteristics. However, in recent decades, this taxonomy has been augmented, and in certain cases superseded, by molecular genetic techniques. The use of marker genes, particularly the 16S locus, for microbiological classification has become widespread in the decades following Woese and Fox's seminal paper (Woese and Fox, 1977). Such molecular techniques have become the gold standard for clinical and academic assessments of intra-uterine infection. From end-point and real-time PCR, to Sanger and next-generation sequencing (NGS) assays, our perspective on intra-uterine infection is sharpening. Much of these developments are down to the 'renaissance [of] the pioneering 16S gene' (Tringe and Hugenholtz, 2008).

The transcribed 16S ribosomal ribonucleic acid (16S rRNA) gene, usually about 1500 base pairs (bp) in length, forms a structural scaffold for the messenger RNA reading, small subunit (30S) of all prokaryotic ribosomes. The gene is present in all bacterial and archaeal genomes. The eukaryotic equivalent subunit, 18S rRNA, has no sequence overlap. As well as being critical to all prokaryotic life, the 16S rRNA gene has become a principle component of the microbiologist's toolkit, owing to its well-characterised structure of conserved and variable genomic regions (Figure 1.4), making it an ideal target for phylogenetic, PCR-based analyses.

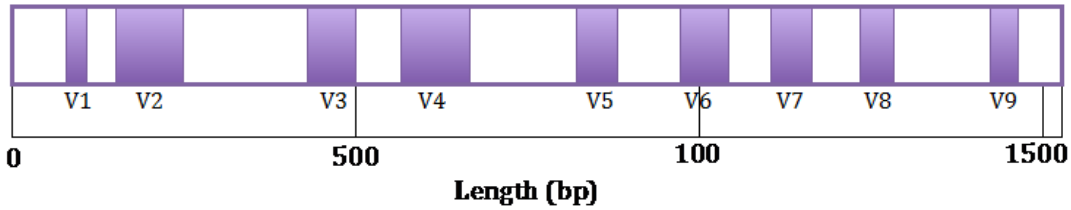


Figure 1.4 - Approximately 1.5 kb of the *E. coli* 16S rRNA gene with its nine variable regions (purple) interspersed between highly conserved sequence.

In genetic studies, primers can be designed with binding sites within either conserved or hypervariable regions, depending on whether species specific or broad-range amplification is required. The design of 'universal' 16S primers that bind to conserved, but span hypervariable regions, enables investigators to describe the full microbiological diversity of a sample in a single experiment. Following amplification,

genetic sequencing of the target region enables characterisation of the microbiological composition of any sample of interest.

The application of targeted 16S amplification translates across a range of molecular biology techniques. End-point 16S PCR describes the presence or absence of specific or broad-range prokaryotic targets, whilst qPCR is used for the quantification of bacterial load within a sample. More recently, the 16S region has been targeted in NGS assays, producing qualitative descriptions of microbial communities. Given the specificity of the locus to prokaryotic genomes, these techniques can be used to investigate bacterial presence in mixed samples, for example investigating bacterial infection within human tissue. However, depending on the region targeted, non-specific binding to phylogenetically related sites on mitochondrial and chloroplast genomes can be an issue.

16S rDNA analysis has been an important component of research into infectious mediated PTB. The work has highlighted the sensitivity of the technique to the very low concentrations of bacterial DNA often present in intra-uterine samples, but also its ability to identify the presence of species not readily cultivable by traditional methods.

1.1.9.1 Limits to the 16S approach

In recent years, the dominance of 16S rDNA PCR in both conventional and NGS microbiological research has been challenged by the advent of whole genome shotgun (WGS) sequencing of entire microbial communities. Metagenomic experiments sequence a random sample of *all* genes that inhabit a particular niche of interest. This enables an investigator to consider the functional and metabolic properties of a microbial community, rather than just its taxonomic and phylogenetic composition. This provides powerful information when considering the functional impact of microbial communities on specific sample types of interest. Furthermore, the reliance on one amplicon in 16S sequencing analyses can lead to amplification biases, and the sequencing technologies used are often limited to relatively short read-lengths. This can lead to issues in the disambiguation of similar organisms, particularly at the species level. Furthermore, results between studies are only reliably comparable if the same variable region is used.

An additional limitation to the use of 16S amplicon sequencing is its reliance on so called 'universal' primers. No primer pair is truly universal in the sense that it will bind with equal efficiency to every prokaryotic 16S sequence in existence. It may in some cases miss out some evolutionary branches entirely, particularly for organisms whose genome has not been yet sequenced (Rosselli et al., 2016). However, *in-silico* tools as the Ribosomal Database Project's (RDP) probe-match algorithm can be used to quantify how 'universal' such sequences are by estimating how many known sequences such pairs will anneal to (Cole et al., 2014). This enables the design of studies that capture as much bacterial variation as possible.

Despite the above criticisms, the 'pioneering' 16S gene remains an essential tool in microbiome research. A broad and highly developed suite of analytical tools is available for the interpretation of 16S specific datasets and it remains the most cost-effective way to carry out microbiome studies. Additionally, read-depths and sequencing lengths are continuing to increase with developments in NGS technologies, increasing the power of these techniques to accurately describe the microbial communities in samples of interest.

1.1.9.2 Using Next Generation Sequencing to infer taxonomy

An incredible increase in sequencing power has been enabled by NGS technological advancements. As a result, high-throughput 16S sequencing has become a commonly used method in microbiological research. Chemical and technological developments in molecular genetics have enabled the production of datasets of tens, if not hundreds, of millions of bacterial amplicons in a single run. The development of specific analysis pipelines and statistical techniques to organise and analyse these data, have run in parallel to such technological progress.

16S sequences must be clustered into smaller groupings in order to be used to calculate diversity estimates or compare taxonomy between outcomes under investigation. There are a number of approaches to classifying and characterising high-throughput 16S amplicon data into a form that can be used to draw meaningful biological conclusions about the sample under study. These are mainly focused on clustering similar sequences into groups known generically as Operational Taxonomic Units (OTUs). Certain OTU-based approaches rely on sequence alignment algorithms such as Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990). Individual or

clustered reads are matched to sequences that have been deposited in taxonomic databases such as GREENGENES, SILVA, or RDP, and are within a specified percentage identity threshold of each other. Alternatively, reads are clustered together, using a variety of different algorithms, again based on percentage similarity but with no reference to an external database. Some techniques combine aspects of the two. In all cases, the intention is to divide sequences into groups defined by some fixed degree of genetic similarity. Such divisions are often interpreted as approximations of widely used taxonomic distinctions. For example, 97% genetic similarity is considered to approximate a species level classification. However, such approximations assume uniform rates of evolution across one locus, such as 16S, which is a scenario that is unlikely to hold in reality. For slowly evolving lineages this uniform threshold may be too relaxed and for those that undergo rapid evolution it will be too stringent (Mahé et al., 2014).

1.1.9.2.1 De novo OTU-picking

De novo OTU picking algorithms cluster reads against each other, without reference to an external database. One of the key benefits to de novo OTU clustering is that all reads are clustered, since they do not have to align to a known sequence. By the same reasoning, de novo OTU strategies enable analysis of samples from the rare or unexplored biosphere, which may include novel, as yet undocumented taxa. This feature of de novo OTU picking can also be problematic if sequencing data has a high error rate or is highly chimeric. In such cases, novel OTUs will be defined by erroneous reads, leading to inflation of measures of diversity within a sample. One of the major limitations of de novo OTU picking is in its speed. The nature of de novo clustering algorithms means that processes cannot be conducted in parallel, making the strategy prohibitively slow for large databases.

1.1.9.2.2 Closed-reference OTU-picking

In closed-reference OTU picking, reads are grouped together based on their similarity to pre-defined cluster centroids from an external reference database. This database may be a public resource such as RDP or SILVA (Pruesse et al., 2007), or an individually curated database of organisms of interest and their appropriate genetic sequence representative. If an input sequence matches nothing in the reference database, within the defined similarity threshold, the sequence is discarded and not used in subsequent

analyses. Clearly this is not appropriate for studies in which samples are taken from poorly characterized niches that may contain novel organisms. The primary advantage of this type of OTU picking is that each read is aligned to a database centroid independently. This process is therefore easily parallelizable, so issues of speed that limit the practicality of de novo strategies are avoided. Furthermore, the quality of sequences in curated databases tends to be high, thereby creating relatively robust OTU clusters.

1.1.9.2.3 Open-reference OTU picking

Open-reference protocols combine the two preceding strategies. First, a closed-reference algorithm is run and all sequences that cluster to a reference database are assigned to a closed-reference OTU. Following this any sequences that failed to cluster via this original route are clustered de novo, with cluster centroids defined internally, rather than with reference to an external database. The computational power, and therefore time, for this process to run lies in between closed and de novo OTU picking, because it consists of aspects that are run both in parallel and serially. This was the strategy used in the sequencing analyses carried out in this PhD.

1.2 Hypothesis and aims

The overarching hypothesis of this thesis was that a significant proportion of sPTB cases within the BBB would be due to bacterial presence in the intra-uterine cavity. Specifically, it was hypothesised that bacterial presence in the placenta would associate with delivery at earlier gestational ages. Differences between preterm and term placental colonisation patterns would be both quantitative and qualitative in nature. Additionally, it was hypothesised that sPTB occurrence would be associated with an up-regulation of pro-inflammatory cytokines within maternal blood, representing aberrant maternal immune responses during pregnancy to insults such as infection.

To investigate these hypotheses, this research project encompassed the following four principle aims:

1) Collate, clean up, and characterise the Baby Bio Bank clinical dataset with a particular focus on the establishment and investigation of the, as yet unstudied, BBB preterm birth cohort (Chapter 4).

2) Test whether there are observable differences in the quantity and type of bacterial species observed within placental tissues according to pregnancy outcome, specifically preterm versus term birth (Chapter 5).

3) Investigate the existence of a so-called 'placental microbiome' across placental tissue from healthy and complicated pregnancies, and examine how any diversions from the norm associate with GA at delivery phenotypes (Chapter 6).

4) Characterise maternal systemic immune profiles in the run up to delivery in both spontaneous preterm and term births, as a potential proxy for an underlying intra-uterine infection (Chapter 7).

Chapter 2: General Materials and Methods

All experimental work within this project was conducted on nested case-control cohorts within the larger BBB collection. The BBB study design, participant recruitment, ethical approval, procurement and access to clinical data, are outlined in Chapter 4.

2.1 Placental sampling and DNA extraction

2.1.1 Equipment and reagents

- RNA*later*® Stabilization Solution (AMBION)
- DNeasy Blood & Tissue Kit (QIAGEN)
- 80% ethanol (HAYMAN)
- TissueLyser LT (QIAGEN)
- FastPrep® Lysing Matrix B Bacteria (Gram +/-) (MPBio)

2.1.2 Sample collection and storage

All placental samples were collected by the hospital's maternity team and dissected by a BBB recruiter, usually within an hour of delivery. For each placenta, 1 cm^3 specimens were excised from four points below the membrane on the chorionic plate (placental parenchyma), close to the umbilical cord entrance. Villous tissue pooled from 6 sites on the maternal side, the basal plate, of the placenta was also collected from a subset of participants, providing opportunity for comparison between tissue at the fetal and maternal side of the uterine cavity. All samples were rinsed in phosphate-buffered saline (PBS) to remove excess maternal blood, placed in barcoded cryogenic tubes along with 5 ml of RNA*later*, and stored at -80°C.

2.1.3 Dissection of placental tissue for extraction

Placental tissue was located and thawed prior to excision of a subsample that was taken forward for experimental analysis. All dissections were carried out in a sterile lamina flow tissue culture hood using sterile scalpels, petri dishes, and appropriate

protective clothing. 20-50 mg of placental tissue was excised from each sample and cut into small pieces to aid tissue and cell lysis during DNA extraction.

2.1.4 DNA extraction from tissue

DNA was extracted from 20-50 mg of placental tissue using the *Qiagen DNeasy Blood and Tissue Kit*, with an additional bead-beating step following chemical lysis. Considerable debate exists around appropriate extraction procedures for microbiome studies that, by definition, require representative extraction of all bacterial species present. Many studies have noted a consistent under-representation of gram-positive bacteria using standard extraction techniques (Rantakokko-Jalava and Jalava, 2002, Harris and Hartley, 2003). Improved protocols aim to more effectively lyse the tough, peptidoglycan rich cell walls of gram-positive species such as *Streptococci*, or the robust and waxy ones of *Mycobacterium* spp., whilst not being too harsh as to degrade those nucleic acids extracted in the process.

Extraction techniques have been shown to impact the relative abundances of species observed, as well as the proportions of eukaryotic to bacterial DNA recovered (Yuan et al., 2012, Sergeant et al., 2012, Willner et al., 2012, Wesolowska-Andersen et al., 2014). A number of strategies using heat, chemical lysis, enzymatic lysis, physical lysis, and, enrichment procedures (Feehery et al., 2013), in combination with standard DNA extraction kits, have been proposed to reduce bias and maximise yield in the extraction of DNA from mixed communities. Optimal conditions will likely vary depending on the types of organisms present, and the types of tissue being analysed. However, physical disruption using silicone beads and agitation has been consistently shown to be an effective method to minimise gram-negative extraction bias (Yuan et al., 2012, de Boer et al., 2010).

In this study, 20-50 mg of tissue was added to 180 µl DNeasy Buffer ATL and 20 µl proteinase K in a 2 ml sterile screw cap tube and vortexed for 10 seconds. Tubes were then incubated at 56°C for 1-3 hours, vortexing 2 to 3 times an hour, until tissue appeared fully homogenized. Contents were then incubated at 95°C for 5 minutes, and spun at 8000 revolutions per minute (rpm) for 1 minute. 200 µl Buffer AL was next added to each vial and pulse vortexed for 15 seconds, followed by incubation for 10 minutes at 70°C. 1/6th of a vial of *MPBio Lysing Matrix B* tubes containing 0.1 mm silica spheres were added to each tube and agitated at 50 Hz for 1 minute, followed by

addition of 200 µl of 100% ethanol. All samples were then centrifuged at 8000 rpm for 1 minute and the lysed contents added to individual *QIAmp Mini* columns and spun again at 8000 rpm for 1 minute. A fresh collection tube was then placed below the spin column and 500 µl of Buffer AW1 added to the column, followed by further centrifugation at 8000 rpm for 1 minute and replacement of the collection tube again. 500 µl Buffer AW2 was then added to the spin column and spun at 14000 rpm for 3 minutes. 200 µl Buffer AE was then added to the column and incubated for 5 minutes to maximise yield, before spinning at 8000 rpm for 1 minute into a sterile, labelled 1.5 ml flip-top tube. All DNA was stored at -20°C until required.

A negative extraction control, in which no tissue was added to extraction reagents and the normal protocol carried out, was produced for every round of extractions. 16S transcripts were subsequently quantified and sequenced to assess bacterial contamination of extraction reagents, the existence of which has been shown by numerous studies to be a significant analytical challenge for microbiome research (Cuiv et al., 2011, Kennedy et al., 2014b, Salter et al., 2014), particularly in samples of low biomass, such as the lungs (Jervis-Bardy et al., 2015) and the placenta (Lauder et al., 2016).

2.2 Quantitative Polymerase Chain Reaction

2.2.1 Equipment and reagents

- NanoDrop ND-1000 Spectrophotometer
- MicroAmp® Fast Optical 96-Well Reaction Plate and Optical Adhesive Covers (Applied Biosystems)
- StepOne Plus™ Real-Time PCR Systems with StepOne Software v2.1
- Custom DNA Oligonucleotides (Sigma-Aldrich)
- *Power* SYBR® Green PCR Master Mix (Applied Biosystems)
- UltraPure™ Diethylpyrocarbonate (DEPC) treated water (Invitrogen)
- Qubit® 2.0 Fluorometer with Qubit dsDNA Broad Range Assay Kit (ThermoFisher Scientific)

2.2.2 Primer design and optimisation

Ultra-pure barcoded primers had already been purchased for amplicon sequencing of the V5-V7 region of the bacterial 16S gene by a collaborating laboratory in the Institute of Child Health department of Infection, Immunity, Inflammation, and Physiological

Medicine. Although it would have been desirable to use the same primers for both quantitative and qualitative analyses, the sequencing target amplicon binds to a 390 bp region of the *E. coli* genome, which is relatively large for use in qPCR. Therefore, a number of primer pairs, most spanning a smaller region of the same V5-V7 region, were compared for use in quantitative experiments. Using the RDP Probe Match resource, an estimate of the number of target species to which the primers would anneal was produced. PCR efficiencies of *E. coli* standard curves (see section 2.2.4), at a variety of primer concentrations, were also recorded. These two factors were used as quantifiable outcomes for comparison and selection of optimal primer pairs (Appendix Table A 1, page 248). All primer pair sequences were also compared to the reference human genome (hg38, December 2013) using the UCSC Genome Browser (Kent et al., 2002) *in-silico* PCR tool to check for non-specific binding. No pair had any known match to the human genome.

Based on these comparisons, primer pair 785F-939R (GGATTAGATACCCBRGTAGTC/CTTGTGCGGGYCCCCGTCAAT), which produces an amplicon around half the size of the sequencing amplicon and showed much greater reaction efficiency, was chosen for use in qPCR experiments. Optimal annealing temperatures were assessed using end-point PCR. 61°C produced the best results. Primer concentrations were compared using qPCR and 0.6 pmol/µl was chosen for use in experimental analyses.

2.2.3 Reaction set-up

qPCR was used to assess and quantify the presence of total bacterial DNA within samples extracted from preterm and term delivered placentas. Experiments were conducted using the *Life Technologies StepOne Plus™ Real-Time PCR Systems* machine. The absolute quantification method is used to analyse the quantity of bacterial template in a sample, with reference to a standard curve of five serially diluted (1:10) *E. coli* DNA samples, with each sample run in triplicate. In addition to placental DNA samples, standards, and extraction negatives, 3 water blanks were included on each plate to check for PCR reagent contamination. Table 2.1 outlines reaction volumes and cycling parameters used in the assay.

Table 2.1- PCR reaction mix and cycling parameters used to amplify 16S bacterial sequences.

Reaction component	Final Concentration	µl / reaction
PCR grade water (Bioline)	-	1.5
Power SYBR Green Master Mix (Life)	-	12.5
Forward primer (785F)	0.6 pmol/µl	0.75
Reverse primer (939R)	0.6 pmol/µl	0.75
25 ng of 2.5 ng/µl template DNA	1 ng/µl	10
Total	-	25

PCR cycling parameters:

Step	Temperature	Time	Number of Cycles
Initial denaturation	95°C	10 min	1
Denaturation	95°C	15 sec	40
Annealing and extension, fluorescence reading	61°C	1 min	↓
Followed by melt curve analysis conducted in increasing increments of 0.5°C			

2.2.4 qPCR analysis

qPCR measures the emission of fluorescence from SYBR® Green molecules. SYBR® Green binds to double stranded DNA molecules as they are exponentially multiplied across successive PCR reaction cycles, producing a characteristic qPCR fluorescence curve (Figure 2.1). These data are used to infer RNA or DNA starting concentrations in samples of interest. The concentration of PCR product in the sample initially increases at an exponential rate between cycles, but as reagents become depleted the process slows to a linear increase and eventually plateaus with very few new DNA molecules being synthesised. It is at the exponential phase of template amplification that a threshold level is chosen, and kept constant between plates. This is the most consistent phase of amplification. Cycle threshold (CT) values for each sample are calculated as the point at which their individual fluorescence curves pass this threshold. All placental DNA samples were run in duplicate and the mean CT was taken for each sample. Duplicate CTs that were over 0.5 CT cycles apart were discarded. qPCR for such samples were repeated. If samples were again over 0.5 CT cycles apart, qPCR data for these samples were discarded altogether.

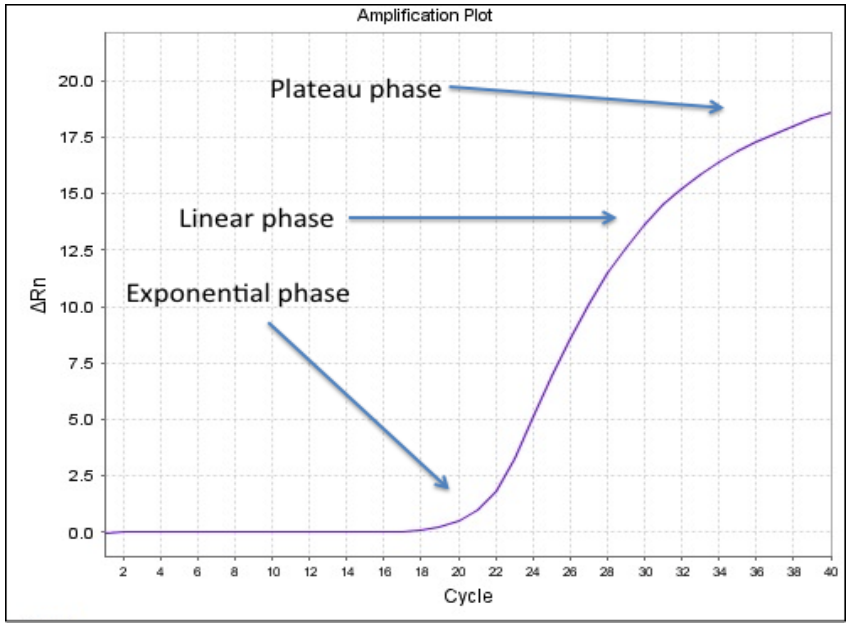


Figure 2.1 - Amplification plot showing the characteristic increase in fluorescence by cycle number that is proportional to the amount dsDNA product in the reaction well.

At the end of each experiment, a melt curve analysis was run to confirm the fidelity of amplification. The presence of one single peak implied the absence of non-specific or primer-dimer amplification, indicating that fluorescence values were representative of only the intended 16S target sequence (Figure 2.2). Results were quality control assessed and analysis conducted using the *StepOne Software 2.1* program.

The absolute quantification method of qPCR measures fluorescence from samples of known concentration in order to calculate the amount of DNA in samples of unknown concentration. By plotting the CT value of standards, in this case *E. coli*, against the $\log_{10}$ of their known concentrations, a 'standard curve' is produced. The mean CT values of each experimental sample can be plotted onto this and their estimated quantity read off (Figure 2.3). Vials with standard concentrations of *E. coli* DNA were kindly provided by the Microbiology, Virology & Infection Control Department at Great Ormond Street Hospital for Children. The first dilution on the curve represents DNA from an estimated 40,000 *E. coli* colony forming units (CFUs). In order to ensure the most accurate quantification, *E. coli* DNA standards were also quantified using a Qubit® 2.0 Fluorometer Broad Range reagent kit each time a new stock was used. Variations in concentrations were recorded and were used to calculate 16S copy number/µl of each *E. coli* standard. 16S copy number for *E. coli* standards were estimated using the relevant *E. coli* genome size (5,231,428 bp), average bp molecular weight (660 gm/mol), Avogadro's constant (6.022E+23 mol^{-1}), and the knowledge that the *E. coli* genome has seven copies of the 16S gene.

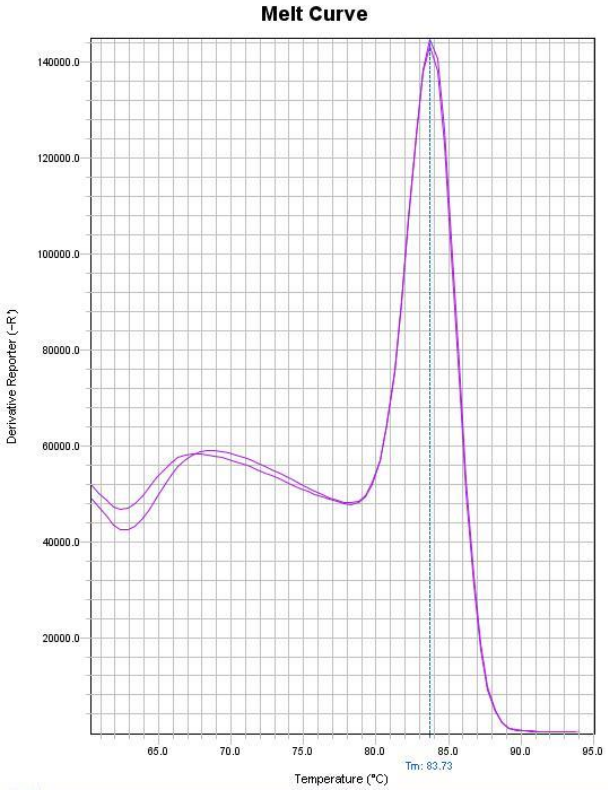


Figure 2.2 – Representative melt-curve peak at around 84°C for two technical replicates indicating fidelity of observed signal to single 16S target sequence.

A standard curve dilution series was run on each plate to minimise the impact of inter-plate variation on sample quantification. The efficiency of each PCR reaction is a metric that describes how closely the reaction is adhering to the expectation of a doubling of PCR product per cycle. Efficiency (E) is calculated by plotting the log of the starting quantity of each standard curve point by their CT value and then inputting the slope of the curve into the equation:

$$E = 10^{-1/\text{slope}}$$

Experiments with efficiencies between 90 and 110% are generally considered to be highly reliable. The mean efficiency for all 35 plates run for this project was 93.9% (standard deviation (SD) = 5.5). Pipetting accuracy is represented by the R^2 value between points.

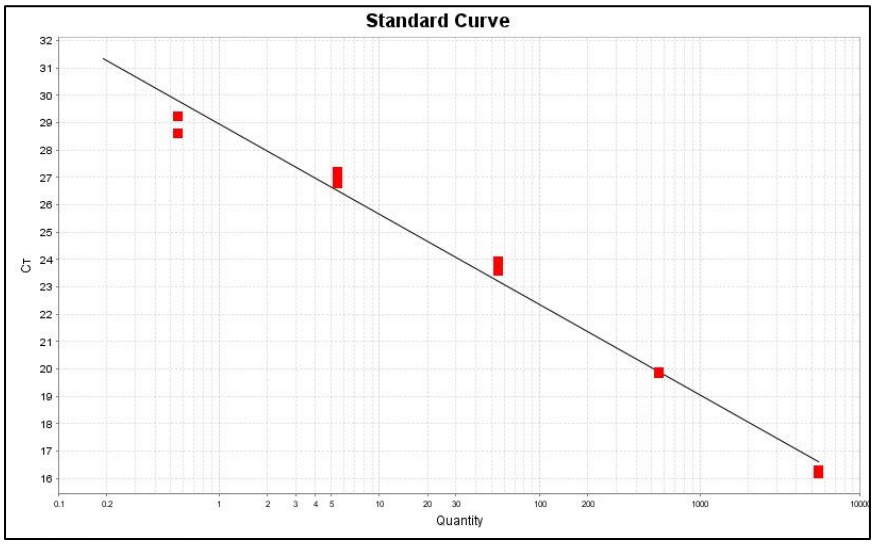


Figure 2.3 – Example of standard curve using 5 (1:10) serially diluted samples of *E. coli* DNA using primers 939R-785F. Efficiency = 100.8%, R^2=0.99, slope =-3.3.

2.3 16S rDNA amplicon high-throughput sequencing

A subset of DNA samples with available qPCR data was taken forward for targeted 16S amplicon sequencing, using a multiplex design, on the Illumina MiSeq platform.

2.3.1 Equipment and reagents

- Veriti® 96-Well Thermal Cycler (ThermoFisher Scientific)
- Thermo-Fast® 96 Non-Skirted (ThermoFisher Scientific)
- MicroAmp® Fast Optical 96-Well Reaction Plate and Optical Adhesive Covers (Applied Biosystems)
- StepOne Plus™ Real-Time PCR Systems with StepOne Software v2.1
- Qubit® 2.0 Fluorometer with Qubit dsDNA High Sensitivity Assay Kit (ThermoFisher Scientific)
- DynaMag™-96 Side Skirted Magnet and DynaMag™-2 Magnet (ThermoFisher Scientific)
- Agilent 2200 TapeStation, D1000 ScreenTape and TapeStation Analysis Software A.01.05 (Agilent Technologies)
- MiSeq™ Next-Generation Sequencer System and MiSeq™ Reagent Kit V2 (500 cycles) (Illumina)
- Moltaq 16S Bacterial DNA-free thermostable DNA polymerase (Molzym)
- Moltaq DNA-free water, PCR-grade (Molzym)
- Moltaq 16S PCR Buffer (Molzym)
- Deoxyribonucleotides (dNTPs) (10mM) (Promega)
- Agencourt AMPure® XP PCR Purification Beads (Beckman CoUlter)
- 80% ethanol (HAYMAN)

- Buffer EB (QIAGEN)
- KAPA Library Quantification Kit for Illumina® platforms (Kapa Biosystems)
- PhiX Control V2 (Illumina)
- Sodium hydroxide (NaOH)

2.3.2 Primer design

Sequencing primers were kindly donated by Dr Ronan Doyle (Department of Infection, Immunity, and Inflammation, Institute of Child Health, UCL). The primers were designed to target the V5-V7 regions of the 16S rRNA gene, 785F: 5′-GGATTAGATACCCBRGTAGTC-3′, 1175R: 5′-ACGTCRTCCCCDCCTTCCTC-3′ (Doyle et al., 2014). The primers were adapted for high-throughput sequencing with the addition of Illumina P5 or P7 adapter sequences, and barcoded dual-index forward and reverse sequences as outlined in a previous publication (Caporaso et al., 2012). The addition of barcode sequences enabled multiplexing of up to 384 samples per run.

2.3.3 Library preparation

Preparation of samples for sequencing of the 16S V5-V7 region involved a 32-cycle end-point PCR reaction. Ultra-pure Taq DNA polymerase (Molzym) was used to minimise the chance of contamination of the sequencing library from bacteria present in PCR reagents. Reaction components and cycling parameters were as shown in Table 2.2.

Table 2.2 – Reaction components and cycling parameters for preparation of library for 16S amplicon sequencing of placental samples.

Reaction component	Final Concentration	µl / reaction
Moltaq PCR grade water	-	13.575
Moltaq PCR Buffer	1X	2.5
dNTPs	180 µM	1.8
Forward primer (785F)	0.4 µM	1
Reverse primer (1175R)	0.4 µM	1
Moltaq DNA polymerase	25mM	0.125
Template DNA	-	5
Total	-	25 µl

PCR cycling parameters:

Step	Temperature	Time	Number of Cycles
Initial denaturation	94°C	3 min	1
Denaturation	94°C	30 sec	32
Primer annealing	60°C	40 sec	↓
Extension	72	90 sec	
Extension	72	10 min	1

2.3.4 Library clean up and quantification

PCR products were double cleaned to remove any primer-dimer and small, non-specific amplicons, using 0.8X Ampure XP beads, according to manufacturer's instructions, and eluted into 20 µl of Buffer EB. Samples were then quantified using the Qubit® 2.0 Fluorometer and dsDNA HS Assay Kit reagents, and pooled in equimolar amounts to the highest possible concentration.

Specificity of the final amplicon pool was assessed using the Agilent 2200 TapeStation and High Sensitivity assay kit, and libraries were re-cleaned using Ampure XP again if necessary (Figure 2.4). Apart from the library for the first run, in which concentration was quantified using the Qubit® and was overestimated, pools were quantified using the *KAPA Library Quantification Kit for Illumina®.* The latter quantifies only those amplicons that are able to hybridise to the flowcell by using Illumina adaptor specific primers and standards. Each qPCR quantified pool was normalised to 2 nM. Finally, up to four PCR plates (386 samples) were pooled into one final pool in preparation for sequencing.

2.3.5 MiSeq set-up

Following library preparation, clean-up and quantification, the pooled library was loaded onto the MiSeq according to Table 2.3.

Table 2.3 – Reaction components, volumes and final concentrations for MiSeq set-up

Reaction component	Final Concentration	µl / reaction
Final 2 nM pooled library	4.5-10 pM	2.25-5
2 M NaOH	-	2.25-5
HT1 MiSeq buffer	-	990-995.5
Denatured PhiX stock	12 pM	100
Index Sequencing Primer (100 µM)	-	3.4
Read 1 Sequencing Primer (100 µM)	-	3.4
Read 2 Sequencing Primer (100 µM)	-	3.4

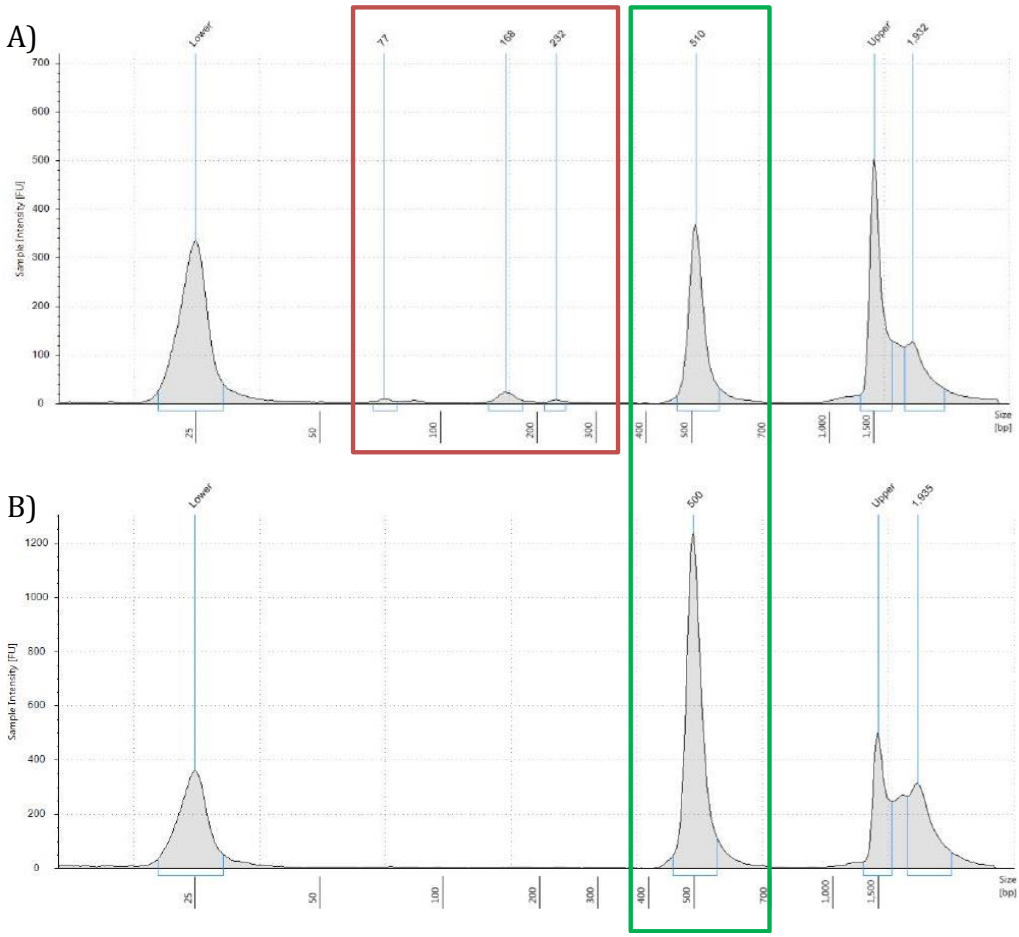


Figure 2.4 - Agilent 2200 TapeStation High Sensitivity trace. Final 16S pooled library of target amplicon at ~502 bp (green box) A) before and B) after final clean-up using Ampure XP beads. A) hows non-specific smaller peaks (red box) no longer present in B) following clean-up

2.25-5 µl of the 2 nM library was combined with an equal amount of 0.2 M NaOH, vortexed briefly and left at room temperature for 5 minutes to denature. Depending on the desired loading concentration, 990-995.5 µl of HT1 buffer from the MiSeq reagent kit was added to the denatured library and mixed well. 100 µl of denatured and thawed 12 pM PhiX stock was then added to 900 µl of this diluted, denatured library.

Index 1 (GAGGAAGGHGGGGAYGACGTTAAAACGTGTT), custom Read 1 (TACCGGGACTTAGGATTAGATACCCBRGTAGTC), and Read 2 (TACCGGGACTTAGGATTAGATACCCBRGTAGTC) sequencing primers were loaded into the cartridge into wells 13, 12, and 14 respectively, and mixed thoroughly. The diluted library was added to well 17. The cartridge was loaded into the MiSeq machine as per manufacturer's protocol for a 500 cycle V2 kit run.

2.3.6 Bioinformatics analyses

Paired-end 250 bp sequenced reads were merged, demultiplexed, quality filtered and assigned to OTUs with taxonomic labels. The UCLUST algorithm (Edgar, 2010) within Quantitative Insights Into Microbial Ecology (QIIME) pipeline was used to pick OTUs at 97% similarity against the Greengenes core reference database version 12.10 (McDonald et al., 2012). Any sequences that failed to match at 97% were clustered de novo using UCLUST. A representative sequence was then chosen for each OTU and this sequence was then aligned to the Greengenes 'Core Set' taxonomic alignment (DeSantis et al., 2006) using PyNAST (Caporaso et al., 2010a). These aligned sequences were used to build a phylogenetic tree using FastTree 2.1.3 (Price et al., 2010). Taxonomy was assigned using the RDP Classifier 2.2 (Wang et al., 2007) and the Greengenes taxonomy reference database, from which an OTU table was constructed. The specific error checking, filtering, and analysis pipeline devised for my samples will be discussed in further detail in 0.

2.4 Multiplex ELISA assay in maternal serum

In order to characterise a selection of the systemic inflammatory responses of BBB mothers in the run-up to delivery, the Bio-Rad 27-plex human cytokine assay was used. This kit assays the concentrations of 27 cytokines simultaneously in each individual biological sample of interest (Table 2.4). The assay is a sandwich enzyme-linked immunosorbent assay (ELISA) protocol in which each of the cytokines, if present in a

sample, will bind to one of the 27 specific capture antibodies provided in the kit that are covalently bound to magnetic beads. These solutions are then subjected to a number of washes to remove any unbound protein, and then combined with a biotinylated detection antibody forming the sandwich complex (Figure 2.5). Finally, streptavidin-phycoerythrin (SA-PE) conjugate is added, which binds to the biotinylated antibodies and serves as the fluorescence indicator when the assay is run.

Table 2.4 – Names and abbreviations of 27 cytokines in the Bio-Rad 27-plex human cytokine kit used to assay maternal serum immunity in Chapter 7.

Cytokine (alternate names in brackets)	
CCL2 (MCP-1)	Chemokine (C-C motif) ligand 2/ monocyte chemoattractant protein 1
CCL3 (MIP-1α)	Chemokine (C-C motif) ligand 3/ macrophage inflammatory protein 1-alpha
CCL4 (MIP-1β)	Chemokine (C-C motif) ligand 4/ macrophage inflammatory protein 1-beta
CCL11	C-C motif chemokine 1/ eosinophil chemotactic protein/ exotaxin-1
CXCL10 (IP-10)	C-X-C motif chemokine 10/ Interferon gamma-induced protein 10
FGF basic	Basic fibroblast growth factor
G-CSF	Granulocyte-colony stimulating factor
GM-CSF	Granulocyte-macrophage colony-stimulating factor
IFN-γ	Interferon gamma
IL1ra	interleukin-1 receptor antagonist
IL1β	Interleukin 1 beta
IL2	Interleukin 2
IL4	Interleukin 4
IL5	Interleukin 5
IL6	Interleukin 6
IL7	Interleukin 7
IL8	Interleukin 8
IL9	Interleukin 9
IL10	Interleukin 10
IL12	Interleukin 12
IL13	Interleukin 13
IL15	Interleukin 15
IL17	Interleukin 17
PDGF-BB	Platelet-Derived Growth Factor-BB
RANTES (CCL5)	Regulated on Activation, Normal T cell Expressed and Secreted/ Chemokine (C-C motif) ligand 5
TNF-α	Tumor necrosis factor alpha
VEGF	Vascular endothelial growth factor

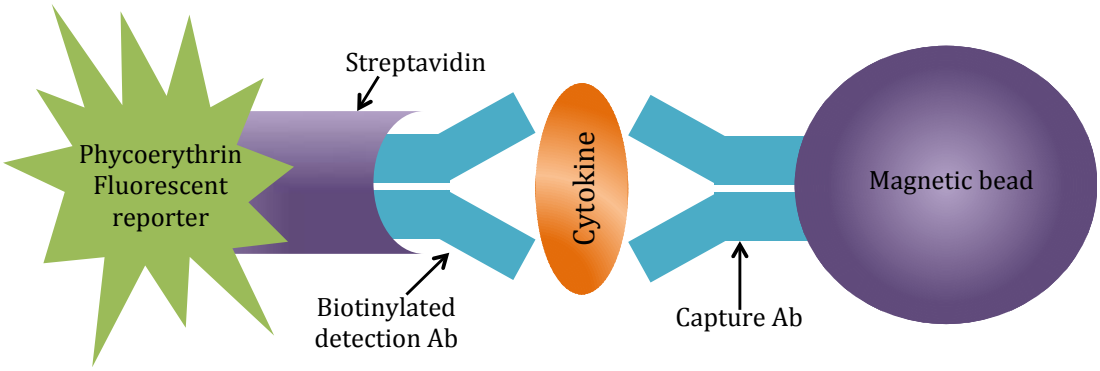


Figure 2.5 – ELISA sandwich complex used in Bio-Rad assay, which results in fluorescence if cytokine of interest is present in sample.

2.4.1 Equipment and reagents

- Bio-Plex® MAGPIX™ Multiplex Reader (Bio-Rad)
- Bio-Plex Pro™ Wash Station (Bio-Rad)
- Microtiter plate shaker
- Standard or sample diluent (Bio-Rad)
- Assay buffer (Bio-Rad)
- Wash buffer (10X) (Bio-Rad)
- Detection antibodies (10X)
- Detection antibody diluent (Bio-Rad)
- SA-PE (100X) (Bio-Rad)
- Coupled magnetic beads (10X) (Bio-Rad)
- Sample or standard
- Sealing tape (Bio-rad)
- 96 well flat bottom plate (Bio-Rad)

2.4.2 Collection and processing of blood for serum

For ease of collection and to encourage participation by minimising the number of blood draws/ hospital visits for mothers, maternal blood samples were collected once during pregnancy, at a time that was convenient to the mother. These were not restricted to a certain time point in gestation. This usually coincided with hospital appointments when blood draws were being taken for clinical purposes. 10 ml of maternal blood was collected according to BBB protocols in clotting tubes (red top) for collection of serum. The appropriate Vacutainers containing the blood samples were spun for 10 minutes at 1200 g and the upper aqueous layer then aliquoted into at least 2 x 2 ml labelled cryovials. Aliquots were stored at -80°C.

2.4.3 Reaction set-up for cytokine experiment

In light of a series of optimisation analyses, outlined in section 3.4, the following protocol was used for the quantification of cytokines in the sera samples of interest. Coupled beads, detection antibodies, and SA-PE were diluted with appropriate quantities of provided diluent. Standards were first suspended in 781 µl of standard diluent, vortexed and left on ice for 30 minutes, and then serially diluted (1:3) to produce a 9-point standard curve. Serum samples were thawed on ice, vortexed and spun down at 10,000 rpm for 10 minutes at 4°C and diluted 1:4 in standard diluent. For each run, along with standards and samples, a duplicate 'blank' sample consisting of 50 µl of standard diluent was also run. This was in order to characterize any background fluorescence that would later be deducted from values observed in experimental samples. The assay was run following the Bio-Rad workflow (Figure 2.6).

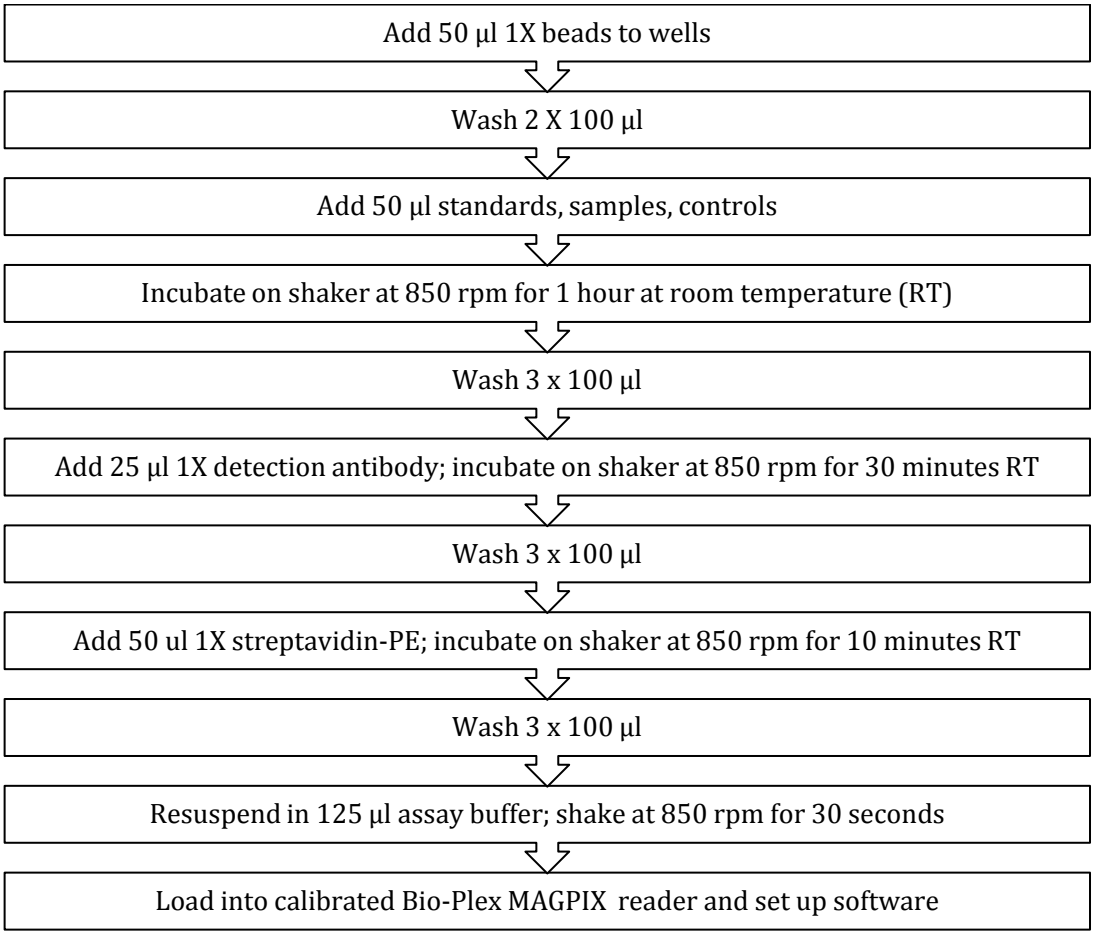


Figure 2.6 – Set-up of Bio-Rad cytokine panel for maternal serum profiling.

2.4.4 Cytokine data analysis

Raw fluorescence values were compared to each plate specific standard curve created by the Bio-Plex Manager™ V6.1 software, to calculate estimated concentrations (pg/ml) for each of the 27 cytokines. Data points above or below the reliable range of the standard curve, or below the background blank value, were flagged as out of range (OOR) by the software. How these missing values were dealt with prior to final analyses is discussed further in Chapter 3.

2.5 Statistical analyses

All statistical analyses were conducted within *RStudio* (version 0.99.484*)* (RStudioTeam, 2015) using the *R* (version 3.3.1 2016-06-21) statistical programming environment (R Core Team 2014). All graphs were created using the package ggplot2 (version 2.1.0) (Wickham, 2009). A significance threshold of 5% ($P<0.05$) was used to define statistical significance, unless otherwise stated. Normality was defined using the Shapiro-Wilk test (Royston, 1982) in R.

In a number of instances throughout this thesis, correlations between technical and biological replicates were quantified using the intra-class correlation coefficient (ICC) calculated within R, using the package 'ICC' (Wolak et al., 2012). The ICC is a value between 0 and 1 that describes how strongly observations within groups, in this case replicates, resemble each other.

Univariate and multiple linear and logistic regression models were constructed using the in-built glm function in R. For certain analyses, a mixed effects model was deemed more appropriate. In such instances, the nlme (version 3.1.128) or lme4 (version 1.1.12) packages were used. Availability of clinical data, interrogation of relevant literature, statistical analyses of my own dataset, and an awareness of the risks of overfitting (Babyak, 2004), formed the basis for my choice of study specific, *a-priori* identified covariates used in my final multiple regression models (Table 2.5). For each study, covariates were combined to form multiple regression models in a single step, forced entry model.

Table 2.5 - Potential confounding variables/ covariates used in multiple regressions for each of the three experimental studies conducted

Study		
qPCR	Sequencing	Cytokines
Delivery method	Delivery method	Maternal ethnicity
Maternal ethnicity	Maternal ethnicity	Batch (fixed or mixed)
Batch	Batch	Smoking
Smoking	Smoking	Maternal BMI*
Maternal BMI*	Maternal BMI*	Maternal age
Recruiting hospital	Recruiting hospital	Parity
Tissue type**	Tissue type**	Days before birth at sampling

* Categorical: underweight, obese, normal ** Where relevant

Chapter 3: Development and Optimisation of Methods

Three core techniques form the basis of the experimental chapters in this PhD: qPCR, NGS, and ELISA. Despite being established techniques, they presented novel challenges for this specific cohort, which required refinement of existing protocols and analytical techniques. Chapters 5 and 6 describe quantitative and qualitative microbial variation within placental samples respectively. DNA extracts from this tissue contain mostly human DNA and, potentially, a very low abundance of bacterial DNA. The majority of microbiome work to date has focused on bacteria rich sites, such as the gut (Wu and Lewis, 2013), in which the ratio between human and prokaryotic DNA is less extreme. Microbiome analyses of sites invariably described as 'sterile', such as the lower respiratory tract (Bassis et al., 2015) and placenta (Aagaard et al., 2014), are rare. It is only recently that methodological and analytical developments specific to these sample types have started to be published (Jervis-Bardy et al., 2015, Glassing et al., 2016, Glassing et al., 2015). Challenges encountered in this project as a result of these sample specific issues were two-fold. The first involved the competing and inhibitory effect of endogenous human DNA on the ability to observe low-level bacteria present in the samples of interest. This impacted both PCR efficiency and sequencing optics. Secondly, it was critical to minimise the impact of contaminating OTUs on study results and interpretation. The influence of such OTUs on the interpretation of qPCR and sequencing outputs has been shown to be larger in low versus rich biomass tissues (Salter et al., 2014).

The third section of this chapter documents the optimisation of the multiplex ELISA assay, the results of which are described in Chapter 7. The sensitivity of cytokine concentrations to sample handling and storage procedures has been reported previously (de Jager et al., 2009). In an attempt to characterise these potential biases, a number of variables were investigated in a preliminary cohort before carrying out analyses on the clinical samples. These investigations were used to establish optimal conditions and analytical models for final experiments.

3.1 Aims

1) Define optimal conditions for quantification of total bacterial DNA from low-biomass placental samples, using V5-V7 16S universal primers.

2) Estimate reproducibility of qPCR assay.

3) Optimise loading concentration of placental DNA samples with low nucleotide diversity for sequencing of 16S amplicons using the Illumina MiSeq platform.

4) Develop a sequencing analysis pipeline to use in conjunction with already available resources, such as QIIME, to account for assay specific error rates and contamination issues in low-biomass samples.

5) Design and conduct an optimisation experiment to investigate impact of freeze-thaw cycling, sample dilution, and batch effect on cytokine profiles in maternal blood samples.

6) Define a protocol for handling the substantial number of missing values produced by the Bio-Rad multiplex ELISA assay.

3.2 qPCR for quantitative analysis of placental infection

The inhibitory effect of too much DNA in a PCR reaction is well known (Schrader et al., 2012). However, this relationship is complicated when working with mixed (e.g. bacterial and human) samples of very low biomass, such as the placenta. In such instances, it is important to minimise the inhibitory effects of background human DNA on the assay, often via dilution of starting template. However, excessive dilution may reduce microbial DNA concentrations below the limit of detection (LOD) of the assay. The following section documents the optimisation of template concentration for use in downstream qPCR experiments. Secondly, reproducibility of the qPCR assay is reported using data from technical replicates.

3.2.1 Methods

See sections 2.1 and 2.2.

3.2.2 Results

3.2.2.1 Optimisation of template concentration for use in qPCR

An inhibitory effect of background human DNA was observed in the placental samples when using V5-V7 16S primers, designed to target only bacterial content. Optimisation analyses were therefore performed in which 3-point standard curves were produced for human DNA samples, spiked with known quantities of *E. coli* DNA. These standard curves were then compared to those produced following amplification of the same concentration of *E. coli* DNA suspended in water. When the total starting amount of either sample type was increased from 1 µl to 2.5 µl in 0.5 µl increments, CT values increased in the mixed samples but reduced, as would be expected, in the pure *E. coli* samples. This observation indicates that although there was more bacterial DNA in either sample type, the effect of more human DNA in the spiked samples was to inhibit the reaction. This effect is demonstrated in Figure 3.1, and a trend was observed for multiple samples of varying starting concentrations.

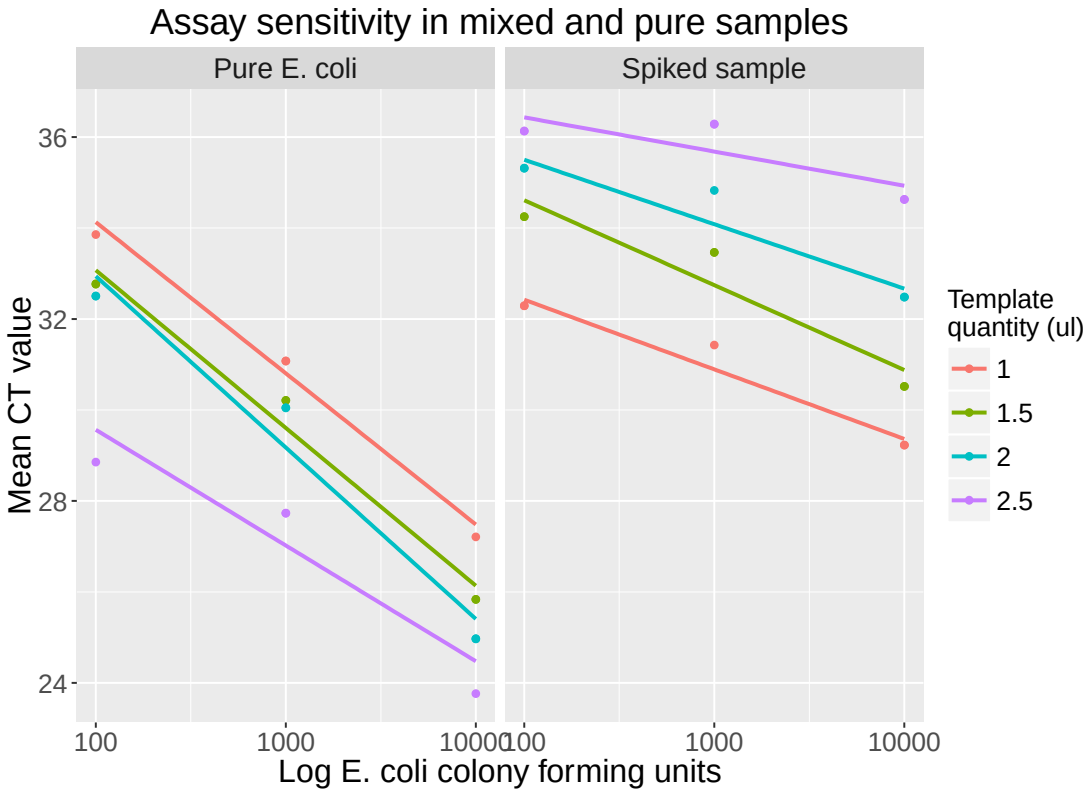


Figure 3.1 – Effect of different amounts of background human DNA on the sensitivity of detection of *E. coli* DNA

To test whether this was just a general effect of too much overall DNA in the sample, regardless of target of interest, the same experiment was conducted using primers to target the human specific GAPDH locus. This did not detect the same inhibitory effect (data not shown). Indeed, there was only a small reduction in CT as the amount of starting template increased, which is to be expected given the increase in starting DNA.

Melt curve analyses of placental DNA samples targeted with 16S primers also indicated that dilution of the starting template increased specificity of target amplification (Figure 3.2). It was next sought to establish the optimum starting amount of total DNA for use in the 16S qPCR experiments. In order to do this, a number of identical samples were run at three different starting concentrations (12.5 ng, 25 ng, and 50 ng). It was clearly visible that the sensitivity of the assay was reduced (i.e. higher CT values were recorded) if templates were diluted too much (12.5 ng) or not enough (50 ng) (Figure 3.3). The 25 ng concentration was therefore chosen as the optimal starting template amount because the lowest median CT values were observed in these samples.

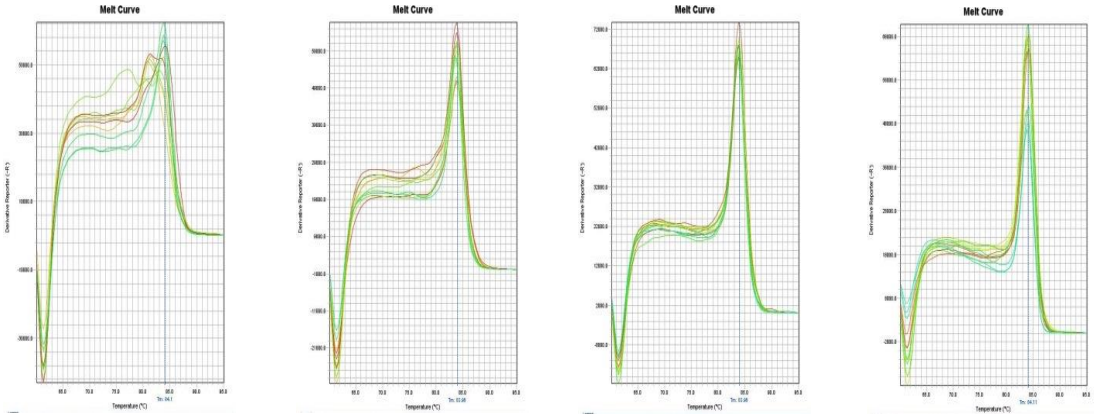


Figure 3.2 – Melt curve analysis of 16S qPCR amplification in 4 serially diluted (1:10) placental DNA samples shows greater specificity with each of the four dilutions (far left = 10^{-1}, far right = 10^{-4})

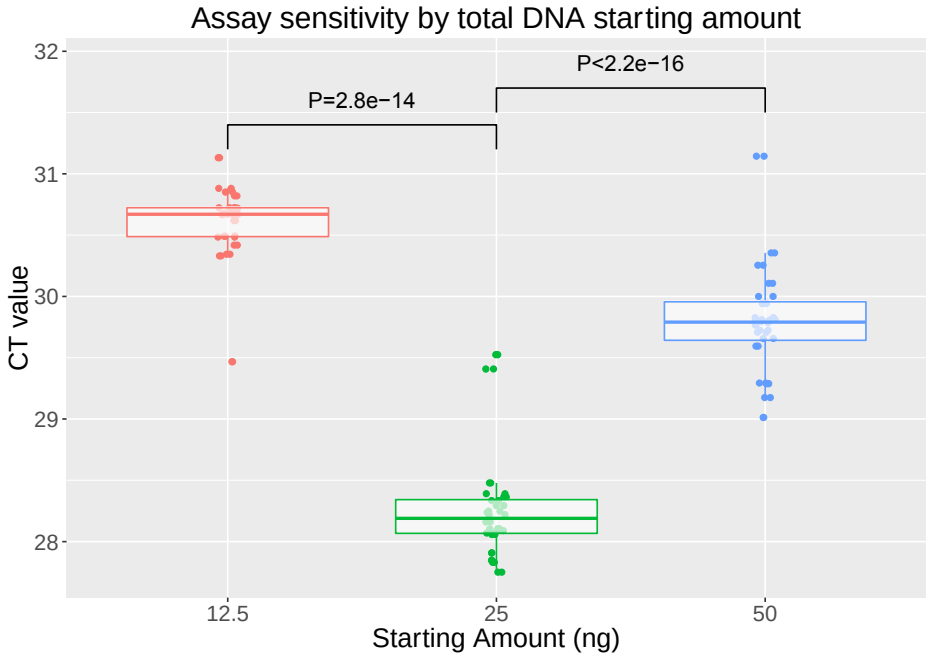


Figure 3.3 – Comparison of assay sensitivity by total DNA starting amount (ng). P-values from Mann-Whitney U test comparison of medians.

3.2.2.2 Reproducibility between technical replicates is high

In total, 26 technical replicates were run across all qPCR plates. These were used to assess the reproducibility of the assay. The correlation between these technical replicates was high (ICC=0.89) (Figure 3.4). Only 39 samples could be run on any one plate, totalling 35 batches. When new *E. coli* stocks were used, concentrations of the estimated 10,000 CFU dilutions were assessed using the Qubit®. This meant any

variation in starting concentration that could affect accuracy of assay quantification was taken into account appropriately (see Appendix Table A 2, page 249 for summary of plate statistics). Despite the potential impact that combining many batches and a variety of standard stocks may have on assay variability, this preliminary analysis provides evidence that the assay was highly reproducible across batches, as long as changes in standards were accounted for.

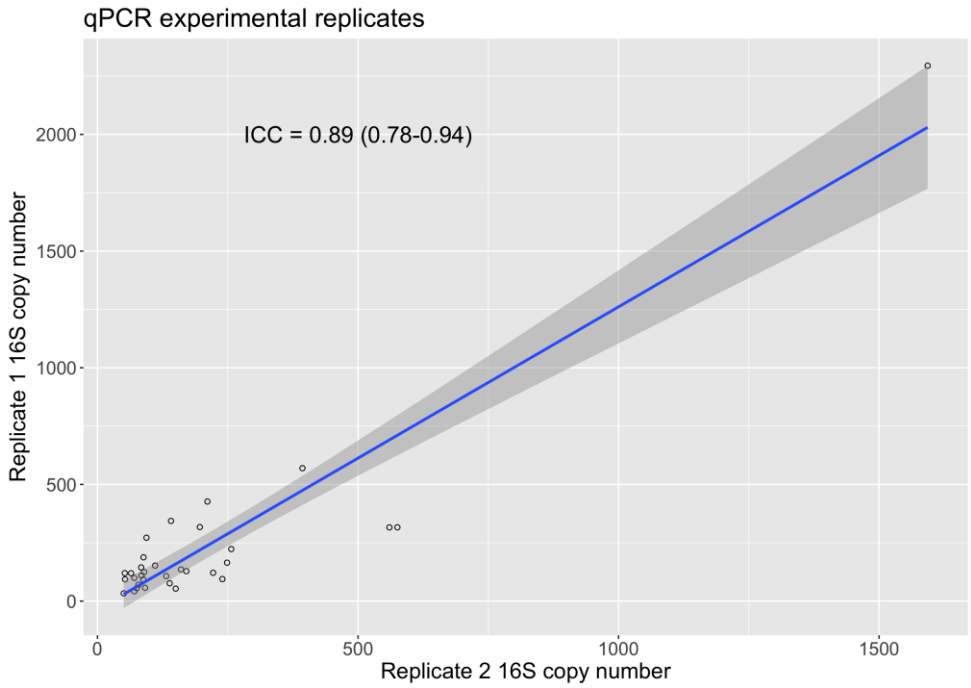


Figure 3.4 - Reproducibility between experimental replicates when 16S quantification takes into account changes in *E. coli* batch concentration. ICC and unadjusted linear regression line with 95% CI show strong relationship between technical replicates.

3.2.3 Discussion

Based on the above investigations, all placental DNA samples were diluted to a uniform concentration such that 25 ng of total starting material was present in each qPCR reaction. This was done in an attempt to maximise assay sensitivity. In addition, each time a new vial of 10,000 CFU *E. coli* standard DNA was collected it was quantified by Qubit®. This value, rather than the estimated CFU count, was used to draw up standard curves for each plate.

3.3 Development of a 16S sequencing protocol and filtering parameters

The low concentration of bacterial DNA in our placental tissue was also a pertinent consideration when conducting the qualitative microbiome analyses reported in Chapter 6. The following section documents the challenges encountered in producing high quality raw data from these particular samples. These were again addressed by adjustment of starting material concentration, generally at the expense of total sample size. In addition, this section will outline the customisation of available computational and analytical tools for analysis of high-throughput microbiome data. These study specific approaches were developed in an attempt to deal with the substantial issue of contamination in the cohort. They were also aimed at increasing statistical power to detect associations between bacterial abundance and outcomes of interest.

3.3.1 Methods

See sections 2.1 and 2.3.

3.3.2 Results

3.3.2.1 Optimisation of loading concentration

Producing high quality sequences from targeted small amplicon libraries can be challenging. This is particularly true of 16S libraries, which have lower per cycle nucleotide diversity in comparison with random shotgun or multi-gene panel libraries. Low diversity libraries are problematic for sequencing by synthesis platforms such as the *Illumina MiSeq* because the two optical channels struggle to find distinct clusters on the images in cycles of limited diversity. For this reason, it is often suggested low-diversity libraries be spiked with a percentage (10-15%) of the highly diverse PhiX genome provided by Illumina. However, even with a 12% spike, my placental libraries were still characterised by very low diversity. In genome libraries prepared using a random fragmentation step, each base would be present at approximately 25%. This would be reflected in a relatively constant, straight line of blue, red, green and black across the percentage base call chart created by *Illumina BaseSpace* software. By contrast, the base-call percentages of my 16S libraries fluctuated extensively across

cycles (Figure 3.5). This was a clear sign of low nucleotide diversity in the samples and proved problematic for the generation of high quality sequencing data.

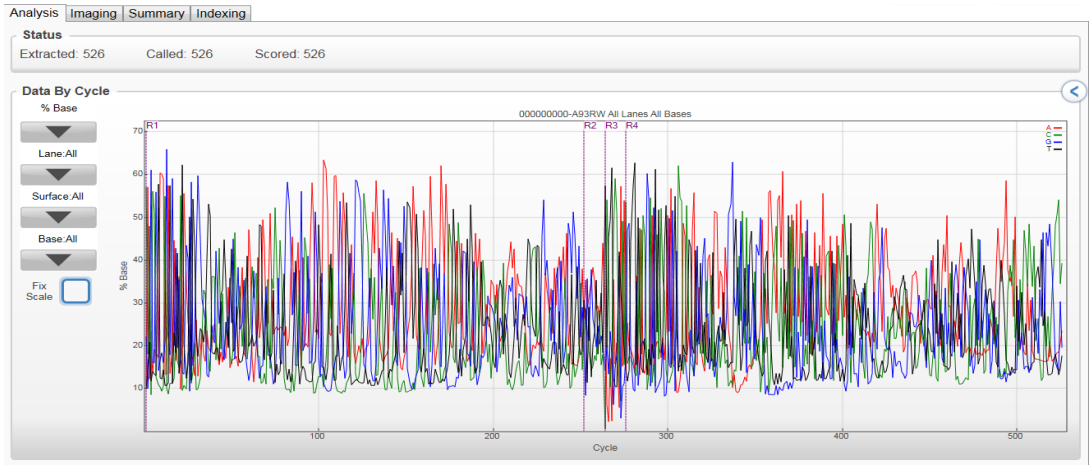


Figure 3.5 - Run 1 percent base calls per cycle. A screen shot from the Illumina sequencing analysis viewer in which percentage base calls (red=A, green=C, blue=G, black=T) are plotted by cycle number.

For reduced diversity libraries, it is generally suggested to aim for a cluster density below the optimum suggested by Illumina for normal libraries (1000-1200 thousand clusters (k)/mm^2) at around 800 k/mm^2. This serves to use the flow cell to a high capacity without compromising on read quality, which deteriorates when the flow cell is overloaded. It proved challenging to identify the optimum loading concentration to achieve this density. The relationship between cluster density and loading library concentration is not strictly linear. It took a number of attempts to identify a reasonable loading concentration to use across runs.

Table 3.1 documents how variation in loading concentration was adjusted across runs in an attempt to improve quality and yield. The first run was of high quality but significantly under-clustered. The second was over-clustered and of much lower quality. Sequence quality is very important for 16S bioinformatics analysis. Therefore, it was decided to concentrate on producing more high quality reads per sample, rather than settling for lower quality data from more samples. Therefore, the final three runs were repeats from the same 386-sample pool. It was not possible to sequence the whole qPCR cohort as had originally been planned. Reads that passed filter (PF) were pooled together before read-merging, quality control and taxonomic/phylogenetic analyses were conducted.

Table 3.1 – MiSeq platform loading concentration affects cluster density and read quality across 4 different runs. % >Q30 is a commonly used threshold that refers to the % total bases on a run with less than 1/1000 chance of being an erroneous call.

Run	Loading concentration	Quantification	% >= Q30	Cluster density (% PF)	Target reads PF (M)
Run 1	10 pM	Qubit*	78.8	224 ± 40 (78.4)	2.65
Run 2	12 pM	qPCR	61.5	1367 ± 70 (23.6)	4.96
Run 3	8 pM	qPCR	70.7	1015 ± 69 (46.6)	8.47
Run 4	4.5 pM	qPCR	71.1	785 ± 97 (63.4)	9.13

* Run 1 was quantified by Qubit, which severely overestimated total concentration.

3.3.2.2 From FastQ to OTU table

Paired-end 250 bp sequenced reads from the four MiSeq runs were merged using Fast Length Adjustment of SHort reads (FLASH) v1.2.11. A minimum overlap of 110 bp and maximum of 160 bp produced reads spanning 342-392 bp, with a median of 370 bp. Before OTU clustering, reads were demultiplexed using their 24 bp Golay barcode and the default options of the QIIME 1.9.1 (Caporaso et al., 2010b) split_library.py command within Python 2.7.11. This command has an option to filter out any reads that fail to meet a user specified Phred quality score (Q score) (Ewing et al., 1998) threshold. Q scores are the most common metric used to assess the accuracy of base-calls from a sequencer, and are used by Illumina to summarise per-cycle accuracy on their MiSeq platform. Q scores are logarithmically related to error probabilities, such that a Q score of 30 (Q30) represents a 0.001 probability of an erroneous base-call.

Analyses from 16S rRNA microbial profiling are highly dependent on the pre-processing steps applied to sequencing output. Removing sequences of poor quality via error-checking algorithms is one such tool that can be used to minimise the presence of erroneous OTUs in the final dataset. Failure to account for reads with high error rates can lead to spurious clustering, inflating diversity estimates. This risks leading to inaccurate biological inferences (Edgar and Flyvbjerg, 2015).

The standard approach for error checking in microbiome pipelines, including QIIME, is to remove reads with an average Q score below a predefined minimum (e.g. 25), and/or truncate and discard reads at the point at which the per nucleotide Q score falls below a certain threshold (e.g. 20) (Bokulich et al., 2013). However, a different strategy has also been proposed that sums the per read Q scores to estimate the expected number of errors in a read, known as Maximum Expected Error (MEE) checking. In this

case a high score indicates low quality. A threshold can then be chosen, normally 1, above which reads are discarded. This approach has been shown to be a more stringent and accurate method of error reduction than traditional average Phred score strategies (Edgar and Flyvbjerg, 2015). Both methods work on the assumption that machine specific error calls are accurate. In the case of *Illumina* runs, this is normally based on output from those reads mapping to the PhiX genome, which are spiked into the library prior to loading.

Given the issues that were encountered regarding low quality sequencing output, it was important to consider such error checking decisions carefully. Truncation on average read and/or per nucleotide Q scores, or filtering with MEE counts were tested via the QIIME split_library.py and VSEARCH fastq_filter commands respectively. Table 3.2 shows that for two different approaches using QIIME, the first being the default <Q25 filtering protocol and the second that discarded reads based on a minimum acceptable Q score of 19, kept a large proportion of reads. However, the number of OTUs was excessively high for such a low biomass library, with almost 30,000 OTUs for both outputs. By contrast, the more conservative MEE approach retained significantly less reads. However, given that a larger proportion of these were high quality, less than a third of the number of OTUs were clustered. This latter approach also better represented the mock communities that were sequenced (data not shown).

Table 3.2 – Reads retained and clustered, and OTUs created following three different error checking strategies.

	Default QIIME	Q19 QIIME	Default QIIME and VSEARCH MEE
Total reads retained	14,294,890	11,277,801	7,103,877
Total OTUs clustered	27074	26507	8644
Total reads clustered (%)	11,235,523 (78.6)	11,161,537 (99.0)	7,099,603 (99.9)

On the basis of these comparisons, the VSEARCH MEE filtering approach was used. High quality reads from all four runs were then pooled. OTU picking, reference alignment, and taxonomic assignment were carried out as outlined in section 2.3.6 using the default QIIME pick_open_reference_otus.py command. Singleton OTUs were removed and chimeric sequences identified and discarded using the uchime_ref command in VSEARCH. The final processed dataset of experimental samples, negative controls, and mocks, thus consisted of 6405 OTUs from 6,440,948 reads. See Figure 3.6 for further details of this process.

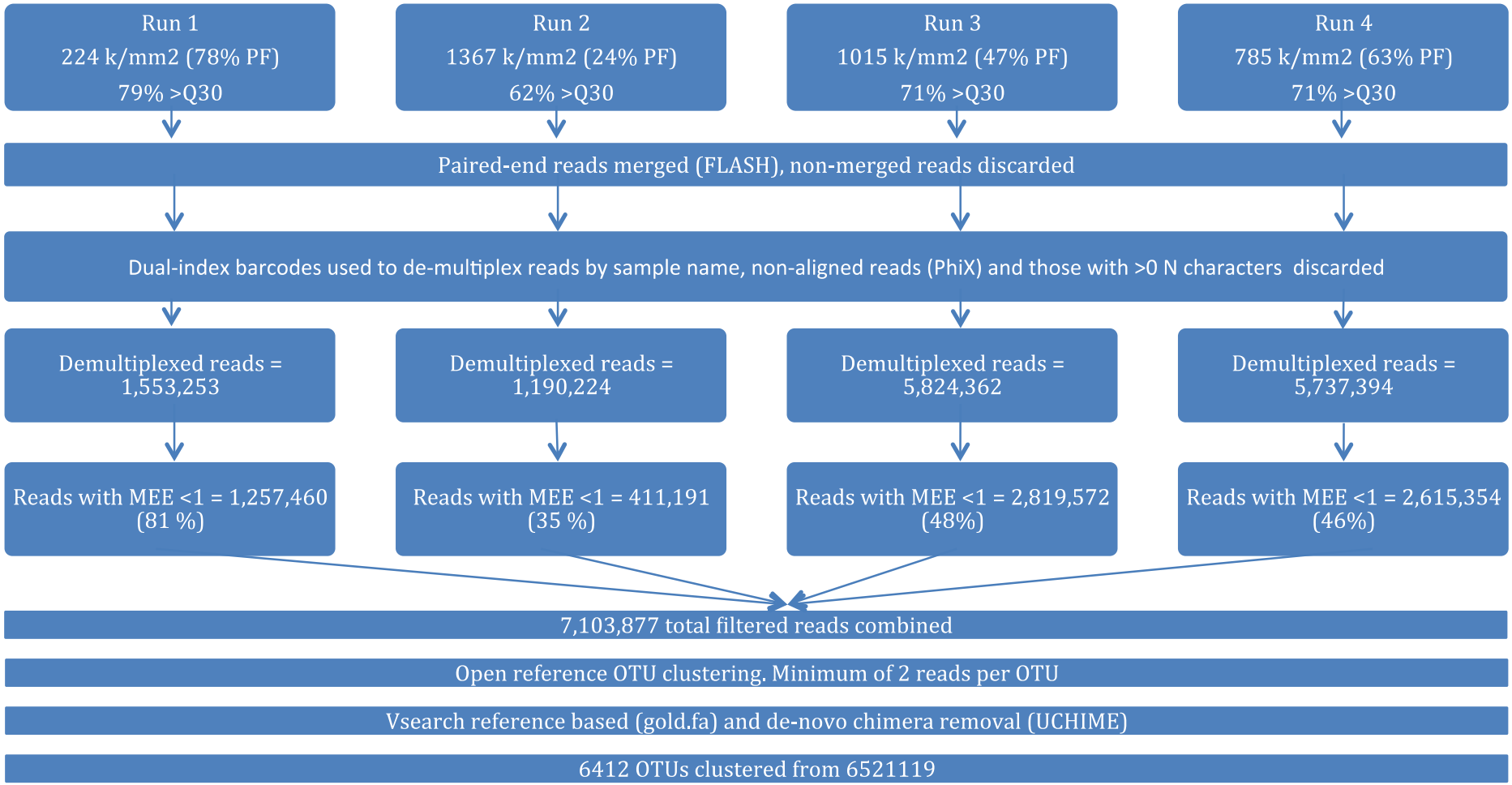


Figure 3.6 – Flow diagram summarising key descriptive statistics of 4 MiSeq runs carried out, and the processes of read merging, initial filtering, pooling and OTU clustering of all resultant 16S sequencing reads that were amplified from placental DNA.

3.3.2.3 Using mock samples to approximate error rate

It is unlikely that initial error checking will successfully identify *all* reads containing either PCR or sequencing error. In cases such as 16S amplicon sequencing, this contributes to the formation of spurious, low-abundance OTUs that would ideally be removed prior to diversity and taxonomic analyses. However, there are no universally agreed thresholds for at what abundance an OTU should be considered low confidence. Indeed, such thresholds are largely study specific. Stringency will depend on how much rare diversity one expects from a sample and the type of analyses that are planned. A common approach is to remove OTUs populated by fewer than 1% of reads (Jervis-Bardy et al., 2015). However, error rates vary considerably across samples and runs. Such arbitrary thresholds do not account for such variability. Furthermore, cut-offs that use average OTU abundance thresholds, result in the removal of rare taxa that may be present at high abundance in only one or two samples. Very low prevalence and/or abundance does not necessarily discount an organism's clinical relevance (Silva et al., 2015).

In this project, study specific error rates were approximated using data from sequenced mock community samples, as an alternative to arbitrary filtering. The intention was to use a filtering threshold to facilitate inclusion of all expected taxa, and maximal exclusion of OTUs that were experimental artefacts. Two independent mock communities were sequenced in the experiments. Mock 1 was obtained through BEI Resources (USA National Institutes of Health (NIH)) as part of the Human Microbiome Project (HMP) (Genomic DNA from Microbial Mock Community B Staggered, Low Concentration, v5.2L, for 16S rRNA Gene Sequencing, HM-783D). This was run on the first plate, with species composition listed in Table 3.3. Mock 2 consisted of genomic DNA samples from 10 species manually combined in units of equal volume, also purchased from BEI Resources (Table 3.4). This mock was run on plates 2, 3, and 4.

Table 3.3 – Species composition of Mock 1 used on sequencing run 1 and OTUs observed matching each expected organism to genus level following initial filtering at 0.002%.

Organism	OTUs identified to genus rank (N)
Staphylococcus aureus, strain TCH1516 *Staphylococcus epidermidis*, FDA strain PCI 1200	53
Escherichia coli, strain K12, substrain MG 1655	26
Streptococcus agalactiae, strain 2603 V/R *Streptococcus mutans*, strain UA159 *Streptococcus pneumoniae*, strain TIGR4	16
Bacillus cereus, strain NRS 248	7
Clostirdium beijerinckii, strain NCIMB 8052	5
Pseudomonas aeruginosa, strain PAO1-LAC	4
Bacteroides vulgatus, strain ATCC 8482	2
Acinetobacter baumannii, strain 5377	1
Actinomyces odontolyticus, strain 1A.21	1
Deinococcus radiodurans, strain R1 (smooth)	1
Enterococcus faecalis, strain OG1RF	1
Helicobacter pylori, strain 26695	1
Lactobacillus gasseri, strain 63AM	1
Listeria monocytogenes, strain EGDe	1
Neisseria meningitidis, strain MC58	1
Propionibacterium acnes, strain KPA171202	1
Rhodobacter sphaeroides, strain ATH 2.4.1	1
Unassigned	1
Other	0
Total	124

Table 3.4 - Species composition of Mock 2 used on sequencing runs 2, 3, and 4 and number of OTUs observed matching each expected organism to genus level following filtering at 0.002%.

Organism	OTUs identified to genus rank (N)
Streptococcus anginosus, strain F0211 *Streptococcus mitis* bv2, strain F0392 *Streptococcus pneumonia*, strain TCH8431	78
Staphylococcus aureus MRSA, strain TCH70	29
Enterococcus facecalis, strain TX1322	16
Peptostreptococcus anaerobius, strain UPII 653-L	12
Lactobacillus crispatus, strain EX849587VC01	10
Escherichia coli, strain MS 110-3	6
Fusobacterium nucleatum polymorphum, strain F0401	4
Treponema denticola, strain F0402	2
Unassigned	12
Other	2
Total	171

Mock samples were subset from the total cohort. The representative sequence of any OTU not identified to the rank of genus, using the QIIME OTU picking pipeline, was compared to the GenBank database using the NCBI Nucleotide BLAST tool (Altschul et al., 1990). Matches with a genus label and at least 97% similarity to the sequence were relabelled in the dataset. Any sequences that failed to match these criteria were recorded as genus 'Unassigned'. Observed OTU taxonomic classifications were compared to the known mock compositions provided by the manufacturers (Table 3.3 and Table 3.4). An error threshold was defined from these data that struck a balance between the maintenance of false negative and false positive calls. Table 3.5 shows how filtering based on the relative abundance of the third encountered erroneous OTU (0.002%), enabled the inclusion of all correct genera across both mocks, along with three, very low abundance, erroneous OTUs. Experimental samples were then filtered based on the results from these analyses. Any OTU not present at a minimum relative abundance of 0.002%, in at least two experimental samples, was removed.

Table 3.5 – Impact of filtering criteria on number of correct and incorrect OTUs identified in mock samples. Depending on the mean abundance chosen for filtering, which was based on abundance at which 1st, 2nd or 3rd erroneous OTU was encountered, accuracy of sequencing results changed.

Filter criteria	Mean abundance	Mock 1		Mock 2	
		N/N correct genera	N incorrect genera	N/N correct genera	N incorrect genera
1st erroneous OTU	2.9 E-04	14 in 17	0	8 in 8	1
2nd erroneous OTU	1.1 E-04	16 in 17	1	8 in 8	2
3rd erroneous OTU	2.2 E-05	17 in 17	1	8 in 8	3

The expected make-up of the mock community samples, along with the number of OTUs matching these expectations to the level of genus, following filtering of OTUs below 0.002%, are shown in Tables 3.3 and 3.4. In Mock 1, all 17 expected organisms were identified to the rank of genus by a total of 124 OTUs, including one OTU that was 'Unassigned' at the genus level. In Mock 2 171 OTUs were retained with at least one OTU mapping to all expected genera, in addition to one erroneous OTU mapping to *Rummelibacillus*; one to *Carnobacterium*; and 12 'Unassigned' OTUs. Figure 3.7 and **Error! Reference source not found.** clearly show that certain genera were present at much higher abundances than others, following filtering. This is, in part, explained by the fact that certain genera (e.g. *Streptococcus*), mapped to more than one species in the

original community. However, such differences could also be the result of stochastic or biased variations in PCR amplification, due to discrepancies in primer annealing efficiencies between different 16S target sequences (Kennedy et al., 2014a). Additionally, 16S copy number can vary between organisms. This may account for some of the variation in total and relative abundances between genera. These observed biases highlight the fact that directly quantitative inferences from sequencing data should be reported with caution. Nevertheless, it is also clear from Figure 3.7 and Figure 3.8 that these differences in abundance were highly reproducible across mock replicates.

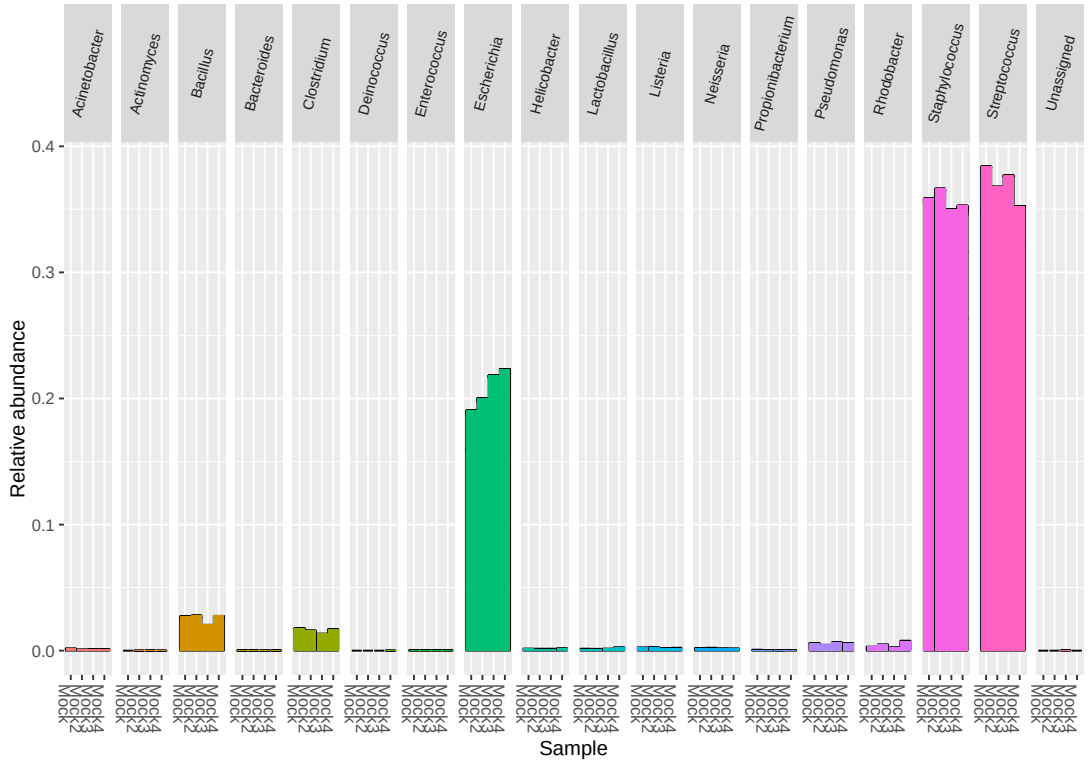


Figure 3.7 – Relative abundances of OTUs remaining in Mock 1 following filtering, which were identified up to level of genus.

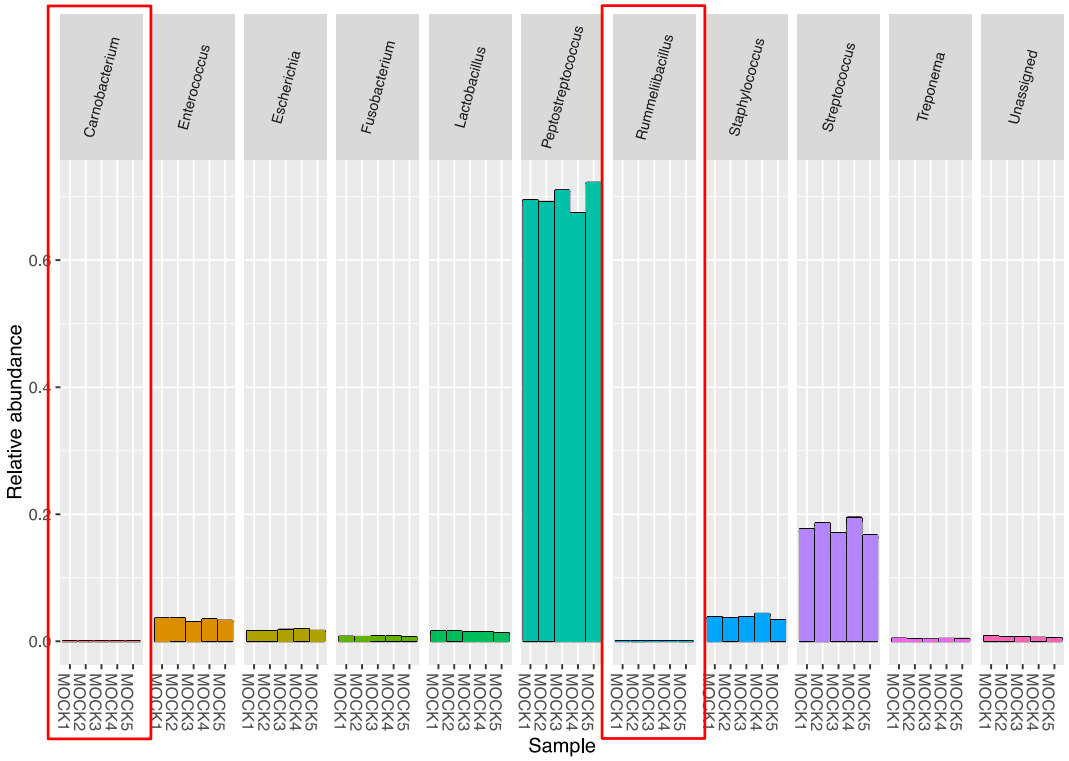


Figure 3.8 – Relative abundances of OTUS remaining in Mock 2 following filtering, which were identified up to level of genus. Red boxes identify genera not present in original community

3.3.2.4 Identification and removal of likely contaminants

Following error checking, negative extractions and PCR blanks were examined for presence of contaminants. Only negatives with >=500 reads were kept (N=16 for extraction negatives, N=3 for PCR negatives). 764 total OTUs were present in the extraction negatives, which decreased to 570 following the removal of singletons. Any OTUs with at least two reads in at least two of the negative extraction samples were considered potential reagent contaminants. A total of 134 OTUs were flagged using these definitions. Although this represented only 2% of the total OTUs identified across all samples, if all these OTUs were removed from experimental samples 50% of reads would have been lost.

In an attempt to mitigate the loss of such a high proportion of reads, a pipeline was devised to ensure that those OTUs that were to be removed were most likely to be true contaminants. One strategy explored the possibility that some of the contaminating OTUs were not reagent contaminants, but instead represented sample-to-control crossover following false index pairings during PCR. This phenomenon has been

reported previously (Rinke et al., 2016, Kircher et al., 2011). To investigate this, the mean relative abundances of OTUs from the 'potentially contaminating' subset were compared between negative extractions and samples. From this it was clear that some of the 'contaminating' genera were in all likelihood originating from experimental samples themselves, rather than extraction kits. For example, whilst 33% of potentially contaminating reads in experimental samples mapped to *Lactobacilli* spp., only 0.34% of negative extraction reads mapped to this genus. *Lactobacilli* are also crucial vaginal organisms. Therefore, it is reasonable to hypothesise that those negative extraction reads mapping to *Lactobacilli* OTUs originated from highly abundant transcripts in experimental samples, rather than from reagent kits. By contrast, most genera previously reported as contaminants, were present at higher or equal abundances in negative extraction, compared to experimental, samples (see Figure 3.9 for such a comparison). Using this rationale, OTUs mapping to *Lactobacilli* spp., *Veillonella* spp., and *Mycoplasma* spp., were considered erroneously flagged contaminants and were not removed from samples. 55% of potentially contaminating reads from experimental samples mapped to these three genera.

The remaining 'potentially contaminating' OTUs were then manually examined for known data on their ecology, and previous evidence for their presence in extraction kits (see Appendix Table A 3 page 250 for details of all genera investigated). Of the 43 genera that the OTUs mapped to, 33 (72%) had been previously reported as reagent contaminants (Glassing et al., 2016). OTUs mapping to genera considered highly likely to be pure contaminants, e.g. *Ralstonia*, a pathogenic plant bacterium, were flagged across all samples. All OTUs mapping to these genera were removed from all experimental samples. Such highly likely contaminants were so defined if they had been previously reported as being present in extraction reagent kits in a recent review by Glassing et al. (2016), and/or they were highly unlikely to be a truly human associated bacteria e.g. *Aureimonas*, a genus of marine bacteria. For other genera the situation was less clear-cut. For example, *Streptococci* spp. are members of the normal human flora, but they have also previously been found in reagent kits. In such instances, only those specific OTUs present in negative extractions were removed from downstream analyses. This retained OTUs in the same genera as these potential contaminants that may have been of clinical relevance, whilst removing OTUs present in extraction negatives. These approaches lead to a substantial reduction in the size of the sample dataset, losing 22% (929,713) of reads.

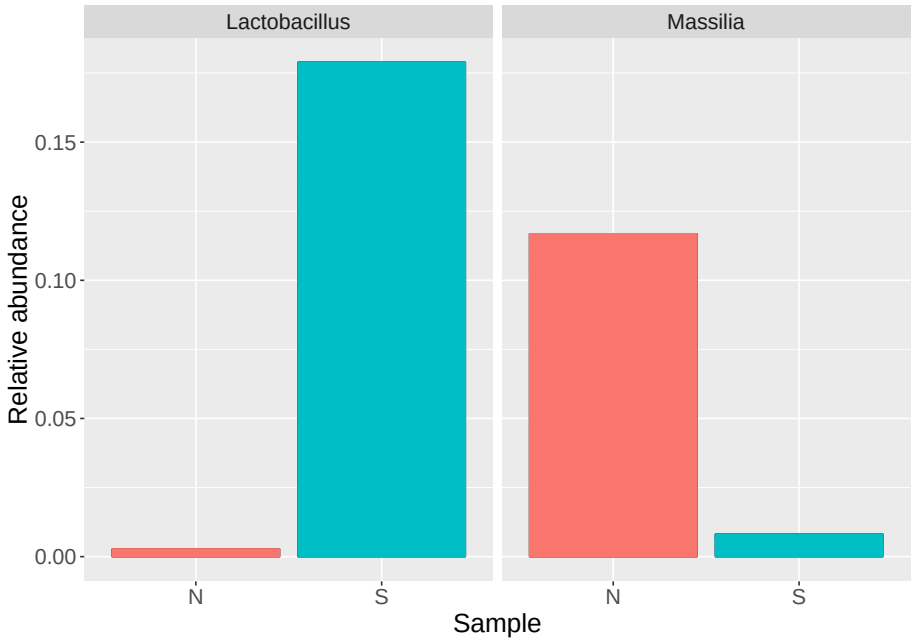


Figure 3.9 - Comparison of mean relative abundances of two potentially contaminating genera between negative extraction (N) and placental DNA samples (S). Substantial differences in abundances likely implicate experimental samples as the original source of OTUs mapping to *Lactobacillus*, and extraction reagents as the source for *Massilia*.

The same process was repeated for PCR water blanks, which in general had far fewer reads than negative extractions (negative extractions median=2637, range=0-37410; PCR negatives median=1012, range=268-2712). Of the three PCR negatives with over 500 reads, only 7 OTUs were present at a minimum depth of 2 reads in at least 2 samples. Using manual inspection and comparison with abundances across samples, it was decided to remove all these OTUs from experimental samples. Additionally, all 15 OTUs that mapped to the *Hymenobacter* genus, an environmental bacterium isolated from various river water and arctic marine sediment samples, were removed (Lee et al., 2017, Kim et al., 2016). After filtering both extraction and PCR negatives according to these steps, a further 3436 reads were removed.

Sample data was thus filtered by contaminant genera, and by OTU abundance, to produce the final filtered dataset. Filtering by OTU abundance removed 3256 low abundant OTUs but lost only 58,152 reads. Conversely, pruning only 291 contaminant OTUs from 40 genera (Table 3.6), removed almost 1 million reads. This highlights the significant sequencing effort that was directed at contaminating OTUs in the dataset. Any sample with under 500 reads was then discarded, because such low sequencing depth is unlikely to be representative of the true diversity of that sample. After all filtering steps were carried out, 272 samples from 189 pregnancies remained, with

2862 OTUs, and 3,162,905 reads. Sample depth ranged substantially from 501 to 239500 reads, with a median of 2160 reads per samples (IQR =1053-7392).

Table 3.6 – Genera matching those OTUs identified in more than one read in at least 2 extraction negatives, and subsequent treatment in experimental samples.

Genera removed entirely	Only contaminant OTUs removed	Not contaminants
Acidovorax	*Acinetobacter*	*Lactobacillus*
Aeromonas	*Corynebacterium*	*Mycoplasma*
Afipia	*Delftia*	*Veillonella*
Arabidopsis	*Dermacoccus*	
Aureimonas	*Enterococcus*	
Brevibacterium	*Escherichia*	
Chryseobacterium	*Finegoldia*	
Cloacibacterium	*Granulicatella*	
Enhydrobacter	*Haemophilus*	
Hymenobacter	*Kocuria*	
Janibacter	*Lactococcus*	
Klebsiella	*Micrococcus*	
Massilia	*Mycoplana*	
Paracoccus	*Neisseria*	
Polynucleobacter	*Peptoniphilus*	
Ralstonia	*Peptostreptococcus*	
Rheinheimera	*Propionibacterium*	
Variovorax	*Pseudomonas*	
	Rothia	
	Sphingomonas	
	Staphylococcus	
	Streptococcus	

3.3.2.5 Technical replicates cluster better following filtering

20 technical replicates of sequenced placental samples remained following filtering as described above. Figure 3.10 shows a heatmap of $\log_{10}$ transformed OTU abundances for all these samples. The columns in the image (sample identifier) are clustered based on the ordination-based PCoA method. The majority of technical replicates clustered next to, or very close-by, each other using this visualisation method. This indicates substantial reproducibility of the technique across runs. Ideally, all amplifications originating from the same DNA extract should cluster side by side within this ordination. Remaining PCR and sequencing bias likely contributes to the small number of replicate amplifications that did not cluster together here. It has been observed that lower template concentrations contribute to greater heterogeneity in sample profiles, which may explain why not all of the technical replicates clustered together (Kennedy et al., 2014a). Reproducibility between replicates, as inferred from their heat-map

clustering tendencies, is substantially improved following filtering, when compared to the same analysis carried out on raw unfiltered reads (data not shown). For all proceeding analyses, OTUs from technical replicates were merged into one larger sample.

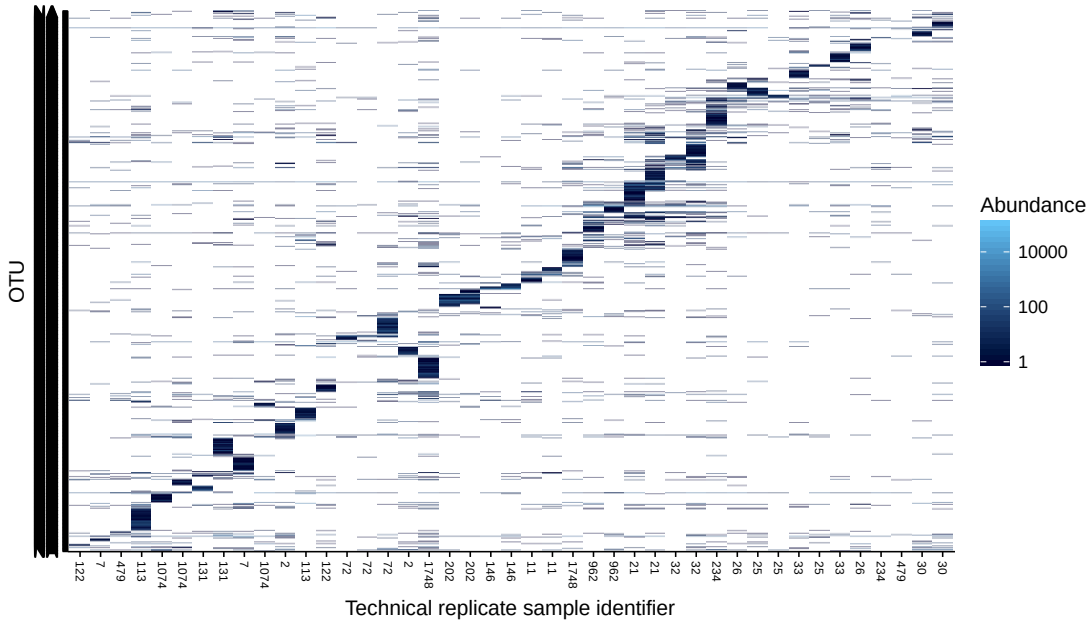


Figure 3.10 – Heatmap of $\log_{10}$ transformed abundances for technical replicates, following filtering procedures. Sample identifiers tend to cluster together, indicating reproducibility of the assay. White = 0 reads present.

3.3.3 Discussion

In this section the development of a study specific filtering pipeline for 16S amplicon sequencing data from placental samples was described. The outlined strategies were aimed at differentiating clinically relevant sequences, amplified from original, endogenous bacterial cells, from contaminating reads. The presence of non-endogenous transcripts in sequencing output may be the result of true bacterial contamination picked up during sample collection and experimental preparation processes. Or such 'contamination' may be a function of PCR and sequencing artefacts. The presence of contaminant OTUs in final datasets will not necessarily impact differential abundance testing, assuming that all samples are equally vulnerable to contamination. However, they will inflate diversity statistics and lead to erroneous biological conclusions when summarising the dataset as a whole. For samples in which little, if any, bacterial DNA was truly present *in-utero*, it is likely that contaminants will form a higher percentage of reads, thus biasing estimates further. This has been

documented in previous studies (Jervis-Bardy et al., 2015). It was critical to remove as many contaminants as possible before conducting experimental tests, whilst also mitigating unnecessary loss of OTUs that in fact originated from endogenous transcripts. To do this a manual filtering process, in which each genus of interest was considered individually, was employed. The development of more specific tools to deal with contamination are needed to improve, simplify, and automate this significant challenge of microbiome research.

The removal of low confidence, low abundance OTUs via error checking and abundance filtering improves the power to detect meaningful biological associations with outcomes of interest. This is because given the size of the dataset, every extra erroneous OTU included in tests will reduce power when associations are adjusted for multiple testing. Strategies to tackle this issue are discussed further in Chapter 6.

3.4 Optimisation of the cytokine assay protocol

During the second year of my PhD I travelled to Virginia Commonwealth University (VCU) in the USA to spend 10 weeks in Gregory Buck's laboratory in the Center for the Study of Biological Complexity. Professor Buck's group specialises in women's health, immunity and urogenital microbiology. In 2014, the team received a prestigious grant as part of the NIH HMP for the establishment of the Multi-Omic Microbiome Study-Pregnancy Initiative (MOMS-PI), a multi-centre collaboration at the forefront of microbiome pregnancy research. The VCU group leads the Vaginal Microbiome Consortium (VMC) as part of MOMS-PI, which incorporates expertise from women's health, pathogenic microbiology, human genetics, epidemiology, and biostatistics.

As part of MOMS-PI, the group at VCU planned to conduct cytokine analyses of cervical swab and maternal plasma samples from their on-going collection of preterm and term pregnancies. Similarly, I planned to investigate maternal immunity and pregnancy outcome using maternal serum samples available from the BBB. During my trip, I designed, conducted and analysed optimisation experiments that would serve to inform both my project and MOMS-PI. The outcome of these optimisation experiments is outlined below.

3.4.1 Methods

See section 2.4.

3.4.1.1 Study design

One blood draw from three independent MOMS-PI participants was diverted from the main study pipeline. Sera and plasma were extracted from each in order to cover the different sample types available in the BBB and MOMS-PI respectively. Samples were assayed using the Bio-Plex Pro Human Cytokine 27-plex Assay from Bio-Rad. The effects on cytokine concentration and assay replicability of the following experimental and sampling variables were considered:

1. Freeze-thaw cycles: each sample was subjected to between one and a total of three instances.

2. Sample dilution: bloods were either diluted 1:2 or 1:4, using standard sample diluent provided by the manufacturer. A previous experiment had shown that running samples neat produced very poor results, so this option was not included.

3. Experimental batch effect: identical samples were run on two separate plates, on consecutive days, by the same investigator. Samples were assigned random positions on either plate, ensuring that any difference between results on these days would represent genuine experimental/stochastic batch effects.

Each combination of the first two variables were assayed in duplicate using blood from three separate participants, such that 36 different experimental samples were assayed (18 plasma and 18 sera) alongside 9 standard dilutions and one assay blank.

3.4.1.2 Statistical analyses

In addition to the general methods outlined in section 2.5, non-parametric, two-sample Wilcoxon rank sum tests were used for comparison of medians. P-values were adjusted for multiple testing by controlling the family-wise error rate using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995) with the p.adjust function in R. Adjusted P-values less than 5% were considered significant. Dendograms were drawn from the output of the hclust hierarchical clustering algorithm within the rafalib R package.

3.4.2 Results

3.4.2.1 Dealing with missing values

Raw fluorescence values from the assay were automatically converted into estimations of concentration (pg/ml), with reference to a 9-point standard curve created for each cytokine, on each plate. Observed standard curve values were considered reliable if they fell within 80-120% of expected values provided by the manufacturer. For certain cytokines, e.g. 1L-17A, not all 9 standard dilutions fulfilled this criteria. Thus, any experimental value that fell outside of this range was deemed to be OOR, either above or below the reliable range of the standard curve. Two other forms of missingness were encountered in the dataset. In the first, observed fluorescence for a sample fell outside of a standard curve in which all points were within the reliable expected range. In this case, the experimental software provided extrapolated values beyond the reliable curve. This extrapolation was considered reliable enough to include in the final dataset.

The final form of missingness occurred when fluorescence values fell below the background control value. These data points were recoded as zero. Any values which fell outside of any standard curves that were unreliable at the extremities were recoded as either half the minimum reliable value, or the maximum reliable value recorded for the cytokine in question.

The numbers of missing values for each of the three categories showed substantial variation between cytokines, and at times, between batches (see Appendix Table A 4, page 252). The pattern of missingness observed between cytokines reflects both the sensitivity of the assay, with respect to the cytokine in question, as well as the reliability of the standard curve on a given plate. For example, across both experimental batches, over 80% of GM-CSF observations were below the background value. This observation either indicates a true absence of GM-CSF within the samples, or insufficient sensitivity of the assay to pick up molecules that were present at very low concentrations. In the latter case, the concentrations are said to fall below the LOD of the assay. By contrast, RANTES showed complete discordance between missingness on plates 1 and 2. The numbers of values that fell above the reliable range of the standard curve were 31 and 0 respectively. This reflects that the top two points in the standard curve from plate 1 fell outside of the acceptable range of the expected value (58% and 130% respectively), whereas only the highest point on plate 2 was out of range for RANTES. As a result of such variation in the types of missingness, the inclusion of cytokines in downstream analyses in experimental assays was considered on a case-by-case basis. Following the handling of missing values, standard curve observations and the assay blank were removed from the dataset. Effects of freeze-thaw, dilution, and experimental batch on assay outcome were then explored.

3.4.2.2 Freeze-thaw

It has been reported that freeze-thaw cycles can have a substantial effect on cytokine concentrations, across a range of different sample types (de Jager et al., 2009). Contrary to expectations, in our cohort samples did not obviously cluster together (Figure 3.11), nor did concentrations appear to change (Figure 3.12), across freeze-thaw cycles. Previous studies have shown that certain cytokines are more susceptible to the effect of freeze-thaw cycles than others. However, in this study no cytokine on the panel showed a statistically significant difference in overall distribution across the three exposures using a one-way analysis of variance (ANOVA) test to compare means (data not

shown). In the majority of cases, cytokine concentrations remained stable across freeze-thaw cycles for identical samples. In cases in which fluctuations were observed (e.g. VEGF), no clear pattern of change was identified using regression models (data not shown).

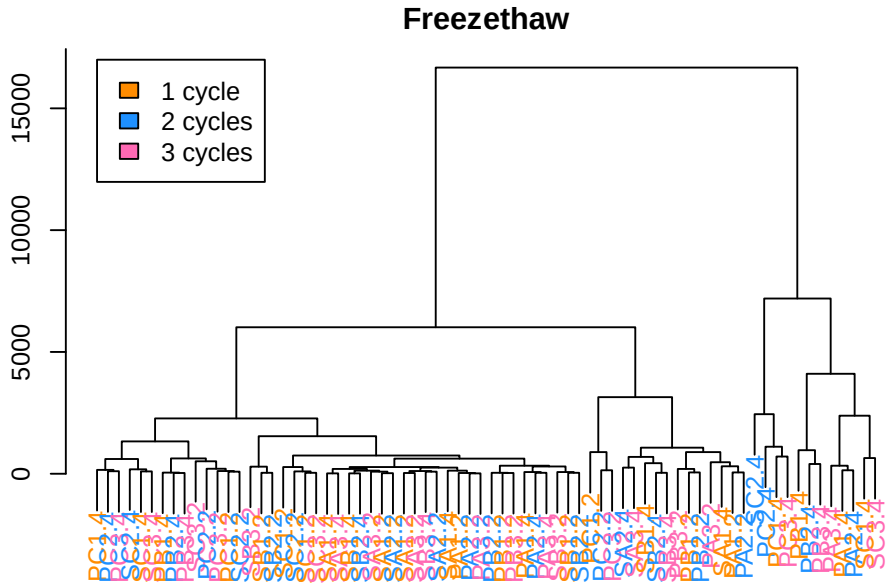


Figure 3.11 – Dendogram produced from samples ordered by hierarchal clustering, using he rafalib package in R. Each leaf represents a sample assayed in optimisation experiments and is coloured according to number of freeze-thaw cycles it was subjected to.

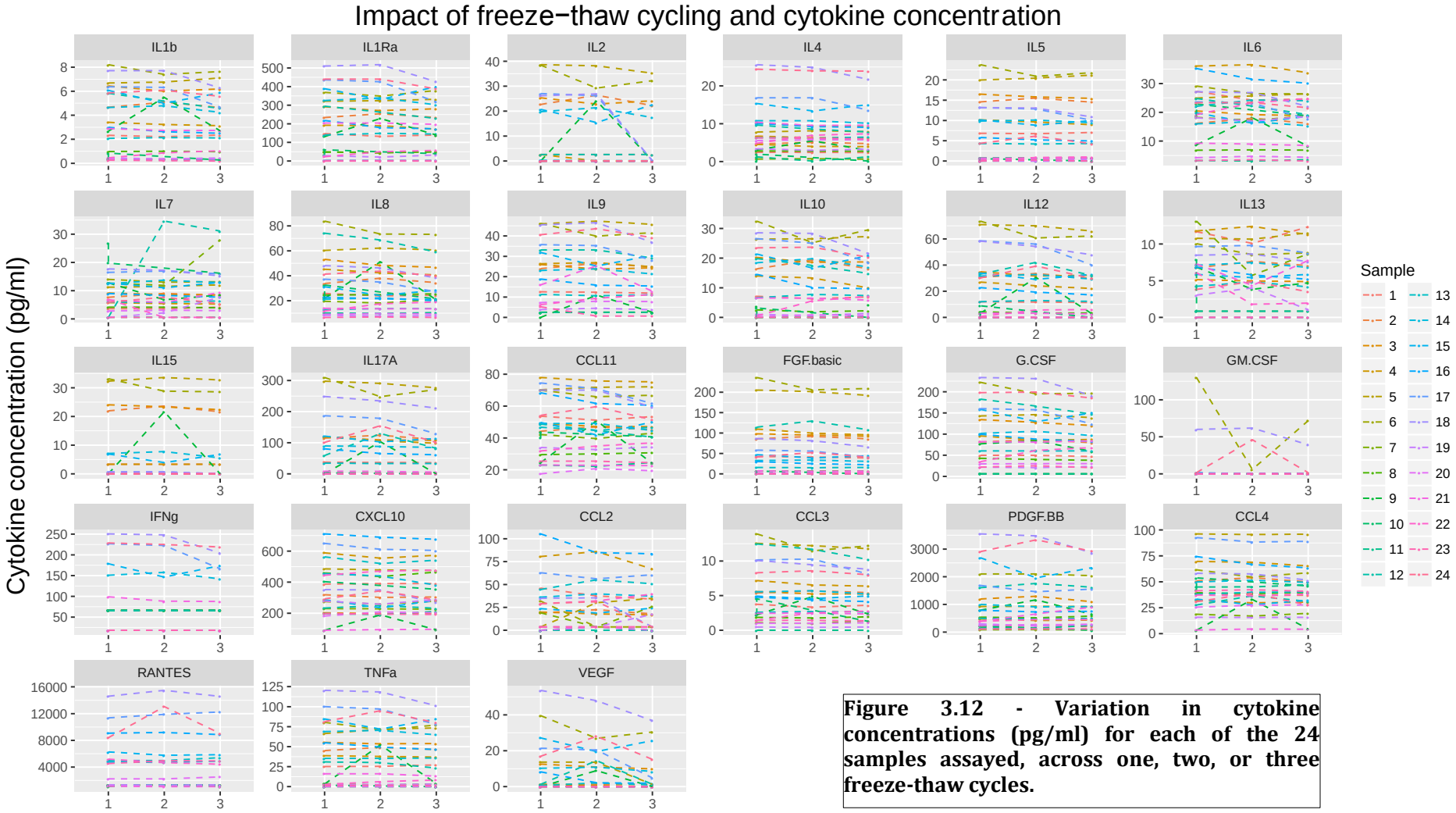


Figure 3.12 - Variation in cytokine concentrations (pg/ml) for each of the 24 samples assayed, across one, two, or three freeze-thaw cycles.

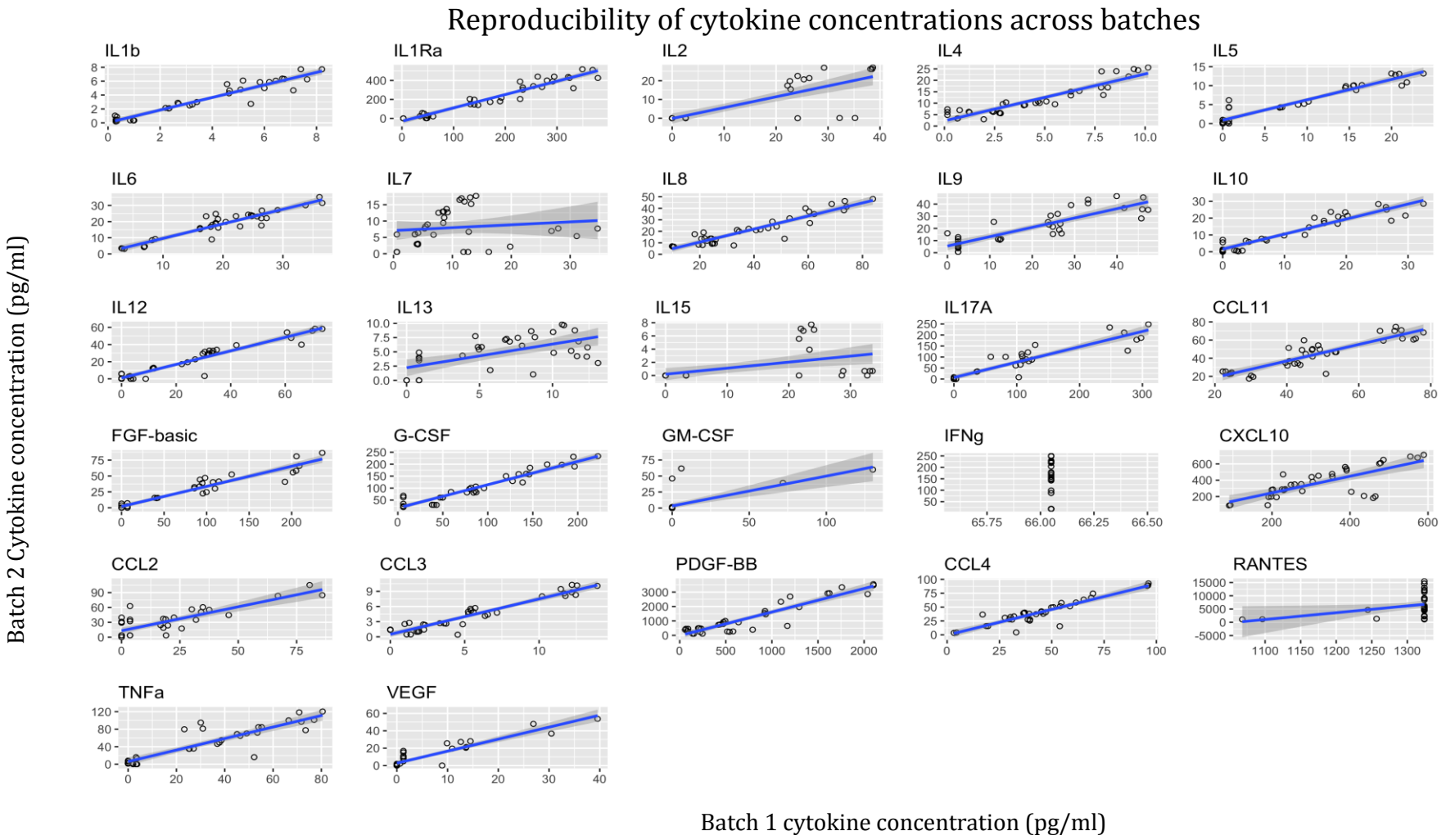


Figure 3.13 – Plot showing relationship between concentrations of technical replicates run on two different batches for each of the 27 assayed cytokines. Blue line is linear regression line with 95% CI of the estimate shaded either side.

3.4.2.3 Dilution

The Bio-Plex Pro Human Cytokine assay had not been carried out on serum or plasma samples in the VCU lab, so optimal dilution ratios of sample to diluent were investigated, to ensure best downstream results. Cytokine concentrations in identical samples were compared when a 1:4 dilution and 1:2 dilution ratio was used (see Appendix Table A 5, page 252 for full comparison of results). Results supported the use of a 1:4 dilution for both blood types. When a more dilute sample was used, distributions tended to span wider ranges and have higher median and mean values. 5 cytokines had identical means across the two dilutions. In the remaining 22 cytokines, all except IL10 and IL15 had a higher median in the 1:4 group. For 14 out of the 27 cytokines, this difference was significant at the 5% level. This difference remained significant for 9 cytokines after adjustment for multiple testing using FDR. Furthermore, except for IL13, the 1:4 dilution produced results with a larger range. These results indicate increased sensitivity of the assay with a more dilute starting material.

3.4.2.4 Batch effects

Experimental batch effects can be powerful confounders, if not properly accounted for, when analysing experiments run on different days, machines, or by different handling personnel. In order to test their impact, two identical batches were run on the same apparatus, by the same handler, on consecutive days. Batch 1 and Batch 2 contained identical samples with randomly assigned positions within either plate to avoid any bias introduced by plate positioning. An exploration of the effects of batch on experimental outcome demonstrated the robustness of the assay to batch effect for most cytokines, and the reproducibility of results between days (Figure 3.13). In 18 out of 27 an ICC value above 0.7 was observed.

Despite generally high correlations between most cytokines, a number showed unacceptably low reproducibility between batches, e.g. IFN-γ and RANTES. The reasons for this poor reproducibility may not be uniform between cytokines. For example, with RANTES, the first plate produced a variable distribution of concentrations, whereas the second plate was almost entirely made up of dummy maximum values. This was also the case for IFN-γ, in which the ICC value is essentially zero, since all the values in batch 2 were identical dummy maximums. This lack of reproducibility is clearly an issue of

inter-plate standard curve variability. By contrast, for cytokines such as IL7 and IL15, in which correlations were very low, an explanation for the discordance is less clear. It may be that this assay is not well optimised for the reproducible quantification of these two cytokines in blood samples. In downstream analyses, data from these cytokines will be considered cautiously within the context of poor reproducibility in this optimisation assay.

Although there were generally high correlations between replicates, substantial variation was still introduced into the dataset across batches. This is clearly demonstrated by hierarchical clustering of the sample data, in which the tips of the dendogram are almost perfectly distinguished when coloured by batch (Figure 3.14). This observation highlights the critical importance of accounting for batch in experimental analyses.

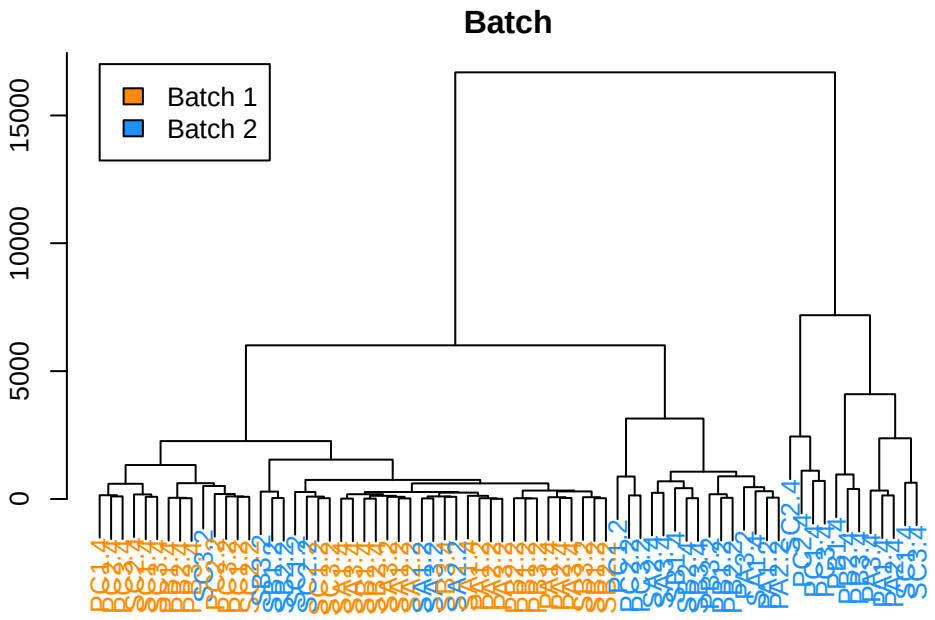


Figure 3.14 - Dendogram produced from samples ordered by hierarchal clustering, using he rafalib package in R. Each leaf represents a sample assayed in optimisation experiments and is coloured according to batch.

3.4.3 Discussion

The sensitivity of ELISA assays to experimental and storage variables such as dilution ratios, batch effects, and freeze-thaw cycle exposure have been previously reported (de Jager et al., 2009). These variables may significantly impact either the true concentration of molecules in a sample, or the ability of the instrument to accurately detect concentrations within a reliable range. These three key variables of interest were tested on a small optimisation cohort. From these data, an optimal protocol using 1:4 diluted samples was established for use with the experimental cohorts from MOMS-PI and the BBB.

Despite the substantial amount of missingness observed in the assay, 19 out of the 27 tested cytokines were reliably quantified, with less than 50% of values missing. This served as a proof of principle for the use of the assay with our experimental samples. Despite substantial impacts of batch on the dataset, correlations were generally high between sampling replicates.

Of particular interest was the marginal impact of freeze-thaw cycles on cytokine concentration. From the results it was reasonable to conclude that any fluctuations across freeze-thaw cycles observed in this assay represented stochastic fluctuations in instrumental measurement, rather than reproducible impacts of the freeze-thaw cycling on cytokine concentration. These observations provided reassurance in cases in which freeze-thaw data has perhaps not been recorded properly. Additionally, this supported the use of technical replicates across different plates that must be refrozen and thawed from one run to another.

Chapter 4: The Baby Bio Bank

All experimental and statistical analyses reported in this thesis were conducted on a subset of clinical data and biological samples taken from the recently completed Baby Bio Bank (BBB). The following chapter summarises the original study design, and clinical and biological sample composition for the entire BBB dataset, including samples not analysed in proceeding experimental chapters. This work formed the initial exploratory component of my PhD, which has been further documented in a recent publication (Leon et al., 2016). All figures taken from Leon et al. (2016), were printed with permission from *Elsevier*.

4.1 Introduction

Common complications in pregnancy affect tens of thousands of women and infants annually. Despite decades of research, and substantial financial investment, the precise molecular aetiology of adverse pregnancy outcomes requires further study. Reproducibility and power in pregnancy research is often limited by the inherent challenges in obtaining biological samples from participants, already heavily burdened with hospital tests and medical visits. The BBB is a large, UK based biobank. It was set up to remove the obstacle of recruitment in research into pregnancy. It is primarily aimed at facilitating research into the environmental and genetic mechanisms underlying common complications of pregnancy. Recruitment to the project began in 2009 and ended in 2015. Pregnant mothers undergoing antenatal care in any of three hospitals across London were invited to participate by trained recruiters. The bank has now finished its sample collection phase. It has amassed over 54,000 biological samples from the mothers, fathers, and infants, relating to 2,515 pregnancies.

Recruitment and sample collection occurred at Queen Charlotte and Chelsea (QC), Chelsea and Westminster (CW), and St Mary's (SM) hospitals. Around 13,000 infants are delivered annually by staff at these centres. Ethical approval for recruitment and collection at these sites was given by the Trent Derby Research Ethics Committee (. Biological sample collection and clinical record retrieval was dependent on the procurement of ethical consent from mothers and, where relevant, fathers, in advance of delivery. Samples are stored at the UCL Institute of Child Health. Duplicates are

stored at SM, Imperial College London. Semi-anonymised patient data were downloaded from clinical records and supplemented by data gathered by the BBB recruitment team. Data are currently stored in a secure database that can be matched to available biological samples.

The four most common complications in pregnancy in the UK are fetal growth restriction (FGR), pre-eclampsia (PE), preterm birth (PTB), and miscarriage. BBB recruitment focused on women with these clinical indications, as well as recurrent miscarriage (RM). A group of healthy women with none of the above complications were recruited as a control comparison group. The underlying aetiologies of these disorders are multifactorial. They involve the interacting and additive impacts of the maternal and fetal genotype; maternal and intra-uterine environmental influences; and clinical interventions. These four common disorders range across a physiological spectrum, covering abnormal implantation to preterm membrane rupture and delivery. It is likely that they share overlapping molecular and physiological mechanisms and pathways.

4.2 Recruitment design

The BBB was principally conceptualised as a genetic and epigenetic resource, with an aim to collect DNA and RNA samples from mothers, infants, and fathers. Plasma, serum, and urine samples were also collected, providing additional opportunities for biomarker research. Recruitment for the BBB followed a targeted prospective cohort design (Figure 4.1). This generally took place either at antenatal or follow-up appointments, often when maternal conditions first presented. Some recruitment also occurred on antenatal and labour wards. Women were only approached prior to delivery. Any consenting mother at any one of the participating units could take part in the project.

A number of strategies were employed to ensure that recruitment to each complication was as high as possible. Women clinically defined as at-risk of particular complications, and flagged to the recruitment team, were particularly targeted. For example, women with a pre-existing hypertensive disorder or history of prior preterm birth, were approached by a recruiter. Nurses and clinicians would also notify recruiters when women threatened complications, such as at the onset of pre-eclampsia, or following preterm membrane rupture. In addition, many women who had normal, uncomplicated

pregnancies were asked to participate in order to establish a BBB healthy control cohort.

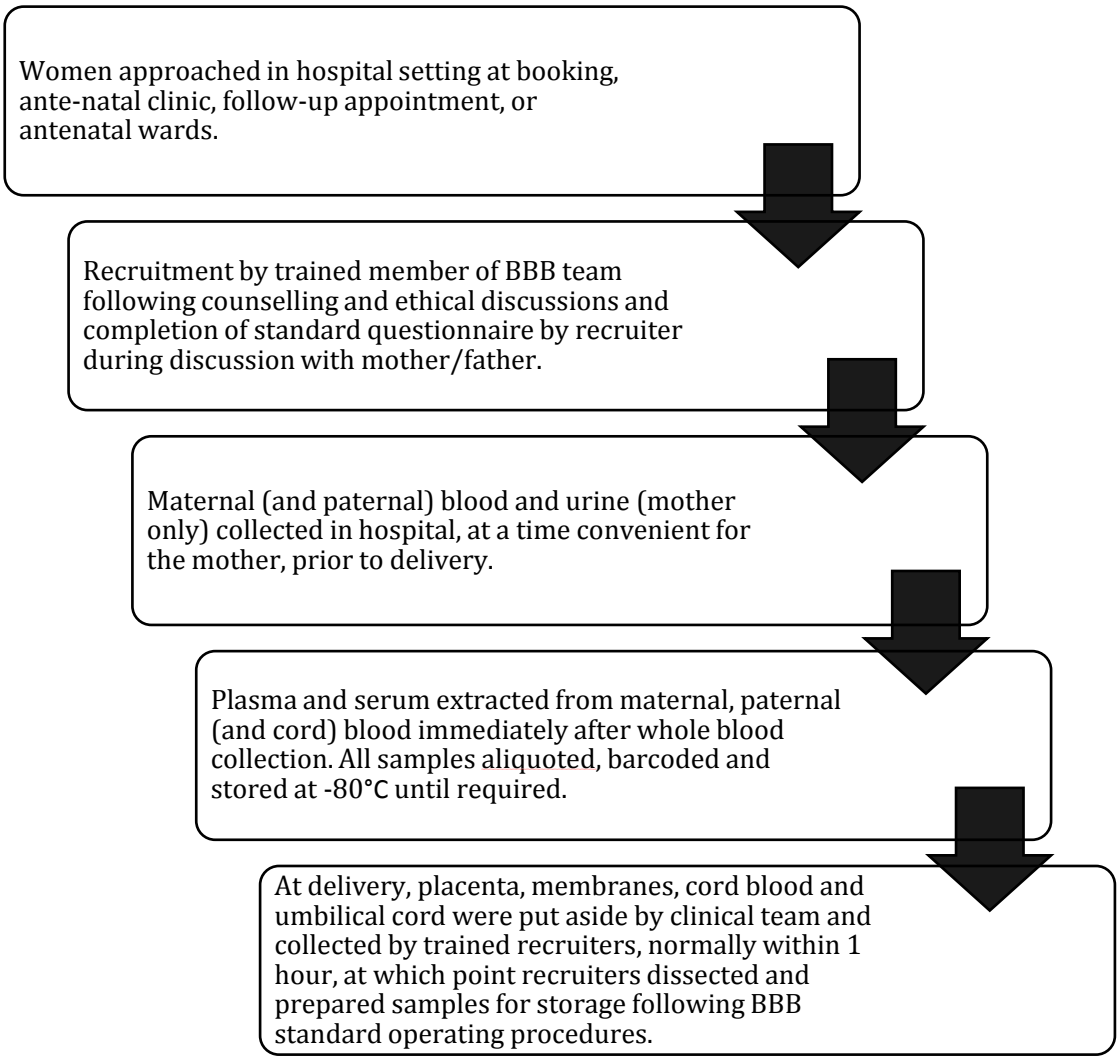


Figure 4.1- Flow-diagram outlining BBB recruitment and sample collection process carried out by trained recruiting staff at 3 London hospitals

The BBB recruitment team consisted of three full-time recruiters who were responsible for counselling and obtaining consent. They also collected and prepared samples around the clock. Recruitment rates were highest at SM (44%), where the project was initiated, 35% were collected at CW, and 21% at QC (Table 4.1).

Table 4.1 – Percentage of participants recruited at each BBB hospital.

Hospital	BBB recruitment N (%)
St Mary's	1107 (44)
Chelsea and Westminster	880 (35)
Queen Charlotte's and Chelsea	528 (21)

The BBB contains data and paired biological samples from 236 FGR, 133 PE, 373 PTB, and 232 RM pregnancies. Over 1,500 'normal' pregnancies with none of the above complications were also collected as a control group. 636 of these were classified as 'perfect' controls with no recorded problems associated with the mothers' health, pregnancy, or delivery. 68% of participating pregnancies also have data and samples from fathers. This is a unique and valuable aspect of the BBB, which, in combination with fetal and maternal tissue, gives it substantial power as a genetic resource.

4.3 Sample collection and storage

The variety of samples that were intended for collection from each pregnancy are presented in Table 4.2. Recruiters focused on securing the collection of 'Trio' sample sets in which tissue, DNA, and RNA were available from mother, father, and baby. This aim was achieved for 1328 pregnancies in total. Each recruitment centre took responsibility for sample receipt and storage. Samples received were matched with clinical phenotype data and barcoded at point of entry. All tissue and blood samples are stored at -80°C until requested, with tissue stored in RNAlater. DNA and RNA are stored at -80°C. These are available for all individuals in which the relevant primary tissue has been collected.

Table 4.2 - Variety of samples collected from participating trios in the BBB.

Sample type
Maternal whole blood/serum/plasma
Paternal whole blood/serum/plasma
Maternal urine
Maternal DNA and RNA (from blood/buccal swab)
Paternal DNA and RNA (from blood/buccal swab)
Placental parenchyma, villous and membrane
Cord blood
Umbilical cord tissue
Baby DNA and RNA (from placenta/buccal swab)

Collecting multiple samples from each placenta increased the total amount of tissue available, as well as facilitating investigations into variation in gene/protein expression in different parts of placenta. . In addition to sampling from the chorionic plate, villous tissue from the maternal side of the placenta, as well as umbilical cord, cord blood, and fetal membrane tissue were also collected, where possible. All tissue was immediately transferred to a vile containing *RNAlater* to minimise degradation of nucleic acids. Occasionally the placenta, cord, or cord blood, were not available, for example, when a baby was delivered at home, in an emergency, or another hospital. In such instances a buccal swab from the baby was requested and DNA extracted from this specimen. Buccal swabs for DNA extraction were also taken from fathers, where blood was unavailable. Maternal blood and urine samples were collected once during pregnancy.

Maternal blood collection ranged from 7 weeks gestation until delivery, with the mean gestational age for collection being 25 weeks. Bloods are available for points collected throughout pregnancy, with peaks at certain times, such as week 12 which coincides with a key antenatal hospital visit for mothers in the UK (Figure 4.2).

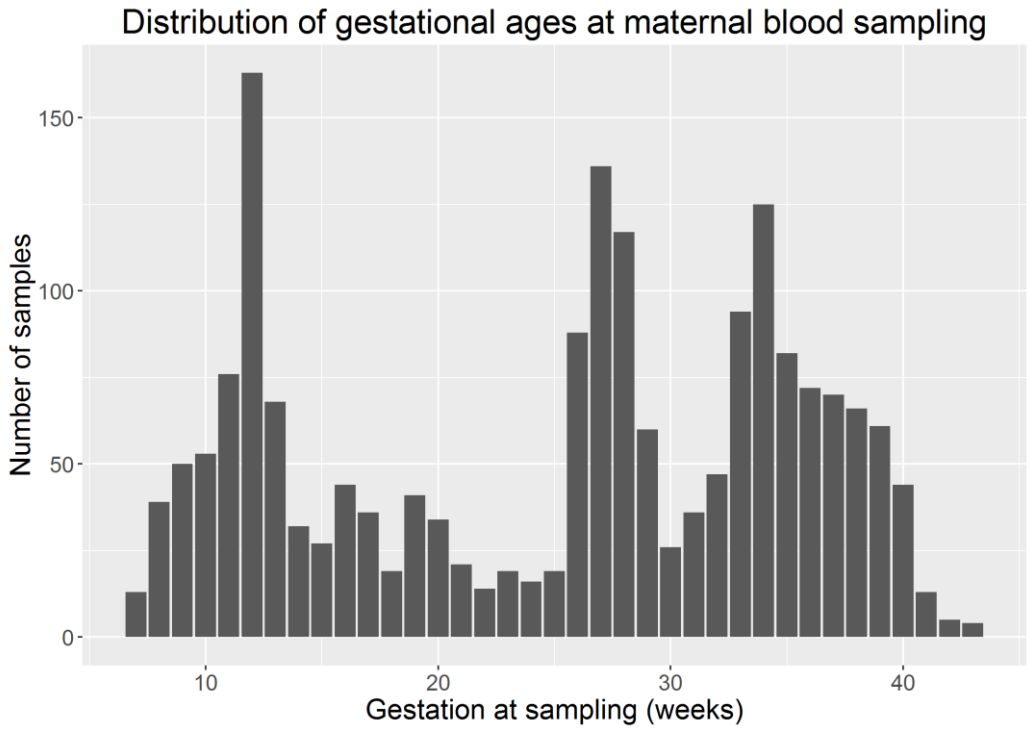


Figure 4.2 - Graph showing distribution of gestational ages at maternal blood sampling within the BBB

4.4 Clinical and demographic characteristics

The BBB clinical database consists of full downloads of electronic maternity records from each hospital, augmented by data collected by recruiters using standardised questionnaires. All entries are available to BBB users on request. Manual surveys via the recruiters, enabled the collection of additional data, which were not available from clinical downloads. This included occupation and information relating to fathers. Furthermore, this enabled the identification of discrepancies in the dataset via comparisons of duplicate categories. Table 4.3 lists a selection of the clinical categories available to users, and their provenance.

Table 4.3- Selection of clinical information available within the BBB clinical database.

Maternal	Paternal	Fetal
Age[1/2]	Age[1]	Gender[2]
BMI[2]	Height[1/3]	Gestation[2]
Height [2/3]	Weight[1/3]	Birthweight[2]
Weight [2/3]	BMI[3]	Placental Weight[3]
Ethnicity[2]	Ethnicity[1]	Head Circumference[2]
Pre-Pregnancy Weight[1]	Diabetes[1]	Birth Length[2]
Parity[1/2]	Hypertension[1]	Apgar Score[2]
Diabetes[1/2]	Smoking[1]	Mode Of Delivery[2]
Hypertension[1/2]	Medication[1]	Congenital abnormalities[2]
Smoking[1/2]	Occupation[1]	
Medication[1/2]		
Occupation[1/2]		
Pregnancy history[1/2]		
Age at menarche[1/2]		
Marital status[1/2]		
Infectious disease[2]		

[1] Information volunteered from patient to recruitment staff
[2] Information gathered from clinical records
[3] Measured by recruiting staff

Originally, the BBB study design stipulated the collection of pregnancies from white European mothers only. This remains the largest ethnic group within the bank. However, given the nature of the underlying multi-ethnic population in London, clinics were also attended by a significant number of mothers of Asian and African origin. Many of these women expressed an interest in participating in the BBB. The BBB team were aware of the significantly enhanced value to the Bank of inclusion of women of all

ethnic backgrounds. As a result, it was decided to recruit from all women, regardless of ethnicity.

Maternal and infant clinical characteristics within the BBB are outlined in Table 4.4. Consensus regarding the definition and groupings used to categorise ethnicity is often lacking amongst researchers. Within the BBB there are two separate columns of data on maternal ethnicity. These are based on two distinct NHS categories, matching those used in the national census.

Table 4.4 - Selection of demographic and clinical characteristics of the BBB cohort.

Variable	Category	Distribution N (%)	Range (SD)	Mean
Maternal Ethnicity	White	1565 (68.2)		
	Black	229 (10.0)		
	Asian	99 (4.3)		
	Other**	401 (17.5)		
Maternal Smoking	Non-smoker	2074 (87.7)		
	Smoker	100 (4.2)		
	Quit within last 12 months	190 (8.1)		
Maternal age at booking	<20	25 (1.3)	14 - 55 (5.48)	32.83
	20-25	187 (9.6)		
	26-30	372 (19.1)		
	31-35	741 (38.0)		
	36-40	478 (24.5)		
	41-45	140 (7.2)		
	>46	6 (0.3)		

Maternal BMI at booking	<18.5 (Underweight)	71 (3.0)	14 - 66 (5.27)	24.87
	18.5-24.9 (Normal weight)	1304 (55.3)		
	25.0-29.9 (Pre-obesity)	630 (26.7)		
	30.0-34.9 (Obesity Class 1)	217 (9.2)		
	35-39.9 (Obesity Class 2)	94 (4.0)		
	Above 40 (Obesity Class 3)	42 (1.8)		
Maternal Parity	Nulliparous	1324 (53.5)	0 - 12 (0.95)	0.67
	Primiparous	797 (32.2)		
	Multiparous	354 (14.3)		
Neonate gender	Male	1182 (52.0)		
	Female	1092 (48.0)		
Number of infants	Singletons	2445 (97.2)		
	Twins	70 (2.7)		
Delivery method	CS	716 (42.2)		
	Vaginal	980 (57.8)		
Birthweight (g)			295 - 5470 (728.92)	3164
GA at birth (weeks) *			20 - 43 (2.81)	38.32

* Assessed from ultra-sound scanning
** Chinese, other Asian, other black, other, and all mixed groups.

4.5 Definitions and verification of outcome data

4.5.1 Preterm birth

373 women were recruited to the BBB who went on to have a preterm pregnancy. 122 were collected as full trios. The statistical work in Chapter 5 was an analysis of the clinical data from all of these preterm pregnancies, whereas the experimental work utilised available placental and maternal serum samples from a subset. Recruiters targeted mothers with a history of preterm birth. In addition, they recruited women presenting at clinics who were threatening preterm delivery, either due to preterm membrane rupture, shortened cervices, preterm contractions, or positive fFN test results. 'Very early preterm' cases, generally classified as below 33 weeks gestation, are a powerful sub-group for research into preterm birth aetiology. However, these are the deliveries that are often most complicated and are therefore the most challenging from which to obtain samples. 96 such valuable cases were recruited for the BBB.

The mean GA at birth for the preterm group was 33 weeks (Table 4.5). The youngest delivery was at just 20 weeks gestation, 3 weeks below the currently defined 'limit of viability' (Seri and Evans, 2008). The distribution of cases across the commonly used sub-phenotypes of PTB are listed in Table 4.5. These categories were defined using data from clinical records relating to membrane, labour, and delivery information. For births in which delivery data were available, 38% were vaginally delivered and 62% were delivered via CS.

Table 4.5 - Number of preterm pregnancies according to GA and mode of delivery within the BBB. Not all fields had complete data.

	Preterm birth data (cumulative N)							
GA at birth (weeks)	Total	sPTB*	sPTL**	PPROM***	Indicated PTB	Missing data on labour/ membrane	CS delivery****	Vaginal delivery
20	1	1	0	1	0	0	0	0
21	1	0	0	1	0	1	0	0
22	2	0	0	1	0	2	0	0
23	4	0	3	4	0	4	0	3
24	6	0	4	5	1	5	0	5
25	16	8	6	8	2	6	2	7
26	15	11	9	10	4	0	6	8
27	30	17	16	15	6	7	9	12
28	33	22	19	20	11	0	17	14
29	41	24	21	21	17	0	24	15
30	61	30	27	25	21	10	31	17
31	80	37	33	31	32	11	40	20
32	96	44	39	36	40	12	49	25
33	132	58	51	47	60	14	73	32
34	186	80	70	64	86	20	94	43
35	252	102	87	85	124	26	121	68
36	373	145	113	123	186	42	172	105

* Spontaneous labour and/or membrane rupture and delivery before 37 weeks
** Labour is spontaneous and delivery occurs before 37 weeks
*** Membrane rupture is spontaneous and resultant delivery occurs before 37 weeks
****Emergency and elective

4.5.2 Pre-eclampsia

PE affects around 2-8% of global pregnancies (WHO, 1988, Roberts et al., 2011, WHO, 2011). It is a dangerous, multi-system disorder. The condition tends to present in later pregnancy. It is estimated to account for 10-15% of all global maternal mortality. Women in developing countries bear the vast majority of this burden (Duley, 2009). PE is also associated with FGR and PTB, thereby impacting fetal morbidity and mortality. It is generally considered to be a pathology linked to poor placentation, with a number of genetic, immunological and lifestyle factors identified as risk factors. However, much of the underlying physiology is still poorly understood. It remains the case that the only treatment for pre-eclampsia is delivery of the placenta: a clinical decision that often results in delivery of premature infants.

Pre-eclampsia in the BBB was diagnosed by obstetricians as new hypertension, i.e. at least two consecutive blood pressure (BP) readings above 140/90 mmHg, or an increase in systolic BP of at least 30 mmHg or diastolic BP of at least 15 mmHg above booking, combined with new proteinuria defined as a protein creatinine ratio of greater than 50, or 24 hour quantitation with a level of greater than 300. Recruitment of PE pregnancies focused on women with a medical history of the complication, or those being monitored or treated for the disorder already. The BBB contains over 700 biological samples from 133 women who developed PE, of which 41 were trios, and 45 delivered preterm. The BBB also contains a number of women with PE associated phenotypes such as pregnancy induced hypertension (PIH) or renal complications (Table 4.6).

Table 4.6 - Number of pregnancies associated with a pre-eclamptic phenotype.

Pre-eclampsia phenotype	BBB N
Pre-eclampsia	133
PIH	57
PIH in previous pregnancy/ies	76
Essential hypertension	98
Cardiac complications	15
Renal complications	16
Thrombosis	8

4.5.3 Fetal growth restriction

FGR describes the condition of a fetus unable to reach its growth potential (ACOG, 2001b, Hillman et al., 2015). Such cases are often recorded as small for gestational age (SGA) at delivery. FGR is associated with perinatal morbidity and mortality and complicates about 5-6% of pregnancies in high-income countries (Brodsky and Christou, 2004, Demetriou et al., 2014, Moore et al., 2015, Harding and Bocking, 2001). FGR is also associated with adult-onset diseases, such as type 2 diabetes, obesity and cardiovascular disease (Forsen et al., 2000, Simmons, 2009, Barker, 1992, Barker et al., 1993).

Fetal growth relies on successful nutrient exchange between mother and baby via the placenta, and is influenced both by environmental and genetic factors. A number of risk factors have been associated with low birthweight: smoking during pregnancy, pre-eclampsia, multiple gestation, and a medical history of low birthweight babies. Other maternal, placental, and fetal factors that correlate with FGR are shown in Table 4.7 (ACOG, 2001a, Harding and Bocking, 2001, Wilkins-Haug et al., 2006, Pollack and Divon, 1992).

Table 4.7 - Maternal, placental and fetal factors that are correlated with FGR.

Maternal factors*	Placental factors*	Fetal factors*
Malnutrition	Placental size	Genetic disorders
Alcohol and substance abuse	Placenta previa	Epigenetic disorders
Medication use	Primary placental disease (e.g. chorioangioma)	Nutrient production
Uterine size	Confined placental mosaicism	Infections
Diminished uterine blood flow	Utilization and production of nutrients	Hormonal factors growth factors
	Diminished umbilical blood flow	
	Hormonal factors	

*Some factors can affect the mother, placenta and fetus

Within the BBB, SGA infants were used as a proxy for FGR during pregnancy. Birthweight for age centiles were defined with reference to the 1996 gender specific four-in-one growth charts produced by the Child Growth Foundation in London. Infants in the BBB below the 10th centile were placed in the FGR group. The BBB recruited 234

such pregnancies, 82 of which were trio sample sets. 112 babies were below the 5th birthweight centile and 50% delivered preterm. The distribution of birthweight for age centiles within the BBB, alongside the mean birthweight for each centile grouping are summarized in Table 4.8. The mean birthweight for this SGA cohort was 1880g.

Table 4.8 - Distribution of centiles and mean birthweights across the BBB.

Centile	BBB N	Mean birthweight (gm)
<5th	112	1555
5th-<10thth	122	2179
10th -< 25th	576	2614
25th-<50th	410	2876
50th-<75th	401	3273
75th-<90th	273	3590
90th-<95th	143	3913
>95th	206	4122
Data not available	272	NA

There is growing interest in pregnancies in which neonates are large for GA (macrosomia). These infants pose particular risks to mothers during delivery. Macrosomia is therefore an important phenotype to study in attempts to minimize maternal morbidity and mortality related to childbirth. Over 300 BBB deliveries were above the 90th birthweight for age centile, and over 200 were above the 95th centile.

4.5.4 Recurrent miscarriage

Miscarriage is defined as the spontaneous loss of a fetus before it has reached viability. This includes losses from conception up to the 24th week of gestation. It is the commonest complication of pregnancy. About 1.5% of couples trying to conceive experience RM (Stirrat, 1990). RM in the BBB was defined as any mother that had experienced three or more consecutive pregnancy losses. Previous studies have shown that the risk of miscarriage increases after each successive pregnancy loss. It can reach 45% after three consecutive losses (Regan et al., 1989).

232 women in the BBB had a clinical diagnosis of RM, of whom 97 were collected as a trio. 538 women within the BBB experienced one or two prior pregnancy losses (Table 4.9). Many of these cases had been attending the internationally renowned centre for

RM based at SM hospital. 770 women in the BBB had had at least one previous miscarriage prior to recruitment, totalling over 1500 miscarriage events.

Table 4.9 - Distribution of number of previous miscarriages within BBB.

Number previous miscarriages	BBB N
0	1666
1	399
2	139
3	121
4	55
5	32
6	11
7	7
8	2
9	3
10	1
Data not available	79

RM has many underlying causes. For example, chromosomal abnormalities (Giorlandino et al., 1998), anatomical uterine defects (Grimbizis et al., 2001), endocrine disorders (Glueck et al., 2000), abnormalities in the immune system, e.g. antiphospholipid syndrome (Rai and Regan, 2006), and thrombophilic disorders, such as those caused by the Factor V Leiden mutation (Bertina et al., 1994).

4.5.5 Other complications

The size of the BBB allowed for the recording of a substantial number of other maternal complications, apart from the four main phenotypes targeted, as detailed in Table 4.10.

Table 4.10 - Summary of extra clinical categories of interest within the BBB.

Maternal clinical category	BBB N
Diabetes	23
Gestational Diabetes	31
Pregnancy from In vitro Fertilization	22
Placenta previa	11
Placental abruption	26
Urinary tract infection	36
Thyroid complications	27
Cardiac complications	29
Epilepsy	48

4.5.6 The BBB and my PhD

The work outlined in this chapter, carried out at the start of my PhD, characterised the biological and clinical composition of the whole BBB dataset, as a resource for future research into complications in pregnancy. I next went on to explore in further detail the pregnancies resulting in the delivery of preterm infants, using a healthy control group as a comparison. This work consisted of explorations of the clinical data at hand, as well as the main experimental work carried out during this thesis: using BBB preterm placental samples to investigate the role of intra-uterine infection in normal and complicated pregnancies. The statistical and experimental results from these investigations are outlined in the following chapters.

Chapter 5: Quantitative Analysis of Placental Bacteria

5.1 Introduction

There exists substantial clinical, molecular, and microbiological evidence that the presence of bacteria in the intra-uterine environment during pregnancy is a risk factor for sPTB. This study was designed to explore this association using placental samples from both term and preterm deliveries collected as part of the BBB. The majority of previous work in this field has focused on infection within the fetal membranes and amniotic fluid (Gravett et al., 1986, Doyle et al., 2014). Therefore, this study contributes novel data from a tissue that is central to pregnancy and fetal outcome, but whose involvement in infectious mediated sPTB remains relatively underexplored. Samples of parenchyma tissue, biopsied from below the membranes on the fetal side of the placental disk, and villous tissue, from the basal plate at the maternal side, were available for analysis. Using the directly quantitative approach of qPCR, total bacterial load within placental samples from a subset of BBB participants was estimated. With these data it was possible to address whether total bacterial load, rather than simply presence or absence of organisms, may be a quantifiable risk factor for sPTB. The availability of these two placental tissue types also enabled the investigation of how bacterial colonisation may vary by placental localisation.

This study also contributes to the growing debate about the existence of a healthy 'placental microbiome' (Aagaard et al., 2014, Aagaard, 2014, Kliman, 2014, Lauder et al., 2016). The placenta has until recently been regarded as sterile, in all but the most extreme cases. However, in the wake of novel research it has been suggested that bacterial colonisation of the placenta may in fact be a feature of normal pregnancy. If a commensal microbiome does inhabit the placenta, this raises novel questions. How do these organisms interact with the developing fetal immune system? Can we characterise any quantitative and/or qualitative transitions from a healthy to a pathological placental microbiome? Do such changes have a causal role in adverse

pregnancy outcomes, such as preterm birth? Some of these questions were addressed in this study, using data on bacterial load from DNA samples, extracted from both term and preterm placenta. In addition, negative extraction controls were sequenced. This allowed the critical issue of how to distinguish between true endogenous signal and contaminant noise, in tissues of low biomass such as the placenta, to be addressed.

The first part of this chapter describes preliminary statistical analyses that were carried out on the BBB preterm and term clinical dataset (section 5.3.1). These analyses were aimed at characterising relationships between potential confounding covariates and gestational age at birth. In order to maximise power to identify any underlying associations between our exposure groups and other clinical characteristics, the whole BBB clinical database was utilised at this stage. Any clinical sample that adhered to the cohort inclusion criteria for experimental analyses (see section 5.2.1), whether or not placental tissue was also available, was included in these exploratory statistical analyses. Experimental results are documented in the latter part of the chapter (section 5.3.6).

The hypothesis for this study was that total bacterial load (16S copy number) would vary between preterm and term phenotypes. The transition of a placenta from a 'healthy' to a 'pathological' state, as inferred from its association with adverse outcomes such as sPTB, would have a quantifiable component. Specifically, the highest load would be observed in sPTB placenta, as this sub-phenotype accounts for the largest proportion of infectious mediated preterm births (Goldenberg et al., 2008).

5.1.1 Aims

1) Characterise statistical associations between potential clinical confounders and key exposures using the BBB clinical dataset.

2) Explore whether mode of delivery of placenta accounts for any observed variation in bacterial load, within the subset of BBB samples taken forward for experimental analyses.

3) Test the hypothesis that placental bacterial load varies by pregnancy outcome, with higher loads associated with sPTB.

4) Examine the presence of bacterial DNA within extraction negative samples in order to investigate the impact of reagent contamination on assay results.

5.2 Materials and methods

5.2.1 Cohort selection

All samples were obtained from the BBB biological database with accompanying clinical data that were cleaned and collated as part of this thesis. All experimental work, including the studies described in chapters 6 and 7, followed a case-control study design. For the qPCR analyses discussed in this chapter, cases were defined as any pregnancy terminating in a delivery under 37 weeks, in which placental tissue was available. Control samples were selected from the BBB healthy control cohort, which consists of any term pregnancy without a recorded complication, i.e. no PE, no IUGR, and <42 weeks gestation. Initially, a random selection of 150 healthy term births delivered between 40-41 weeks and with available placental tissue, were chosen as the control group. However, it was retrospectively decided to include a further 100 randomly selected 37-39 week deliveries to increase the overall cohort size. qPCR analyses were conducted on total DNA from between one and three placental samples per pregnancy. All samples with qPCR data that passed qPCR quality control (QC) were used in final analytical models. Figure 5.1 outlines the process of sample selection from the BBB cohort for use in qPCR experimental work.

5.2.2 Sample collection and DNA extraction

See section 2.1.

5.2.3 Quantitative polymerase chain reaction

See section 2.2.

5.2.4 Optimisation of methodology and reproducibility of assay

See section 3.2.

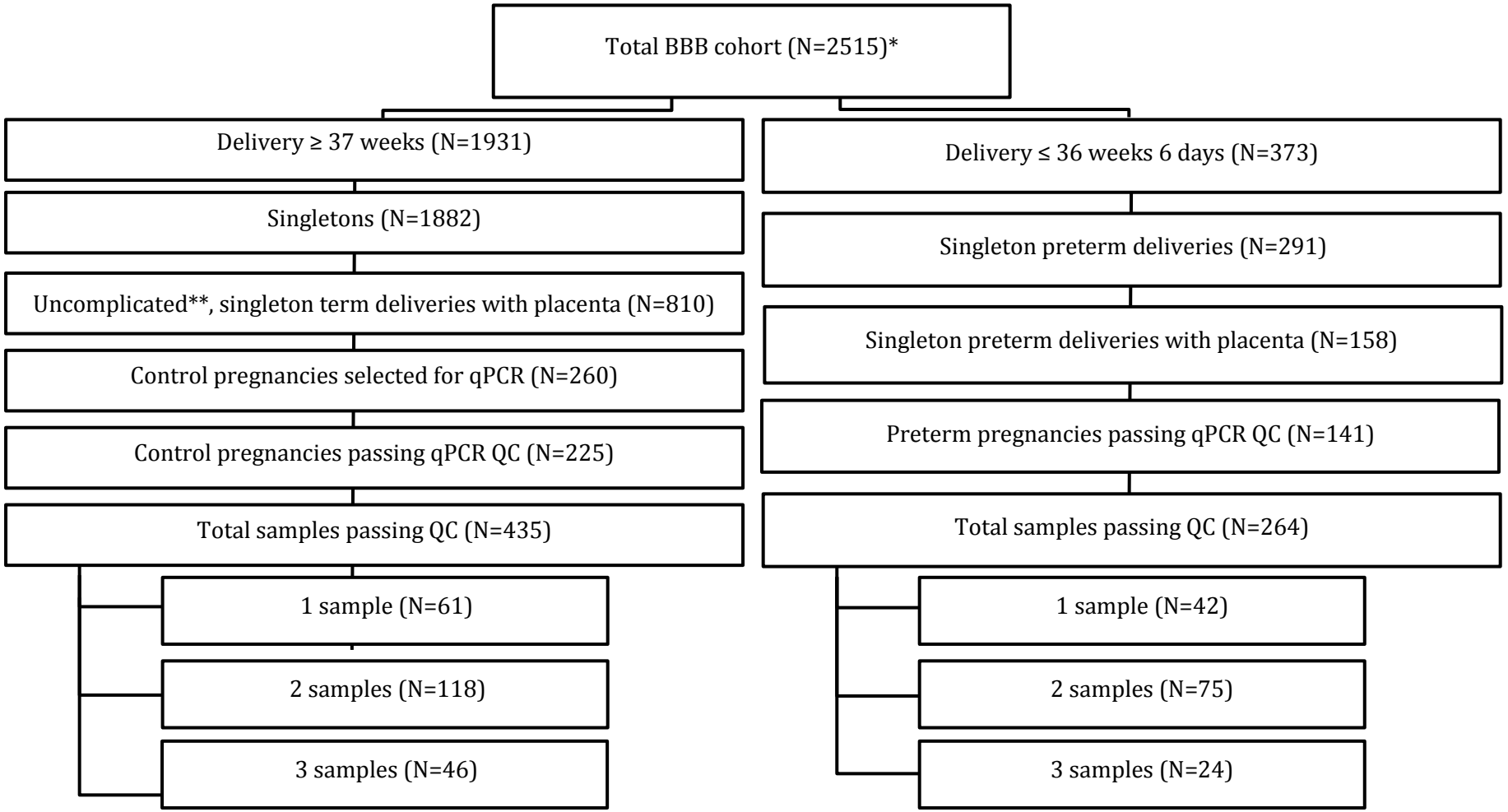


Figure 5.1 – Summary of qPCR experimental cohort construction using samples available within the BBB.

*GA at birth data not available for all deliveries
**No PE, no IUGR, <42 weeks

5.2.5 Statistical analyses

The distribution of qPCR data was highly negatively skewed. Therefore, initial exploratory analyses used non-parametric two-sample Wilcoxon rank sum tests, using the wilcox.test function in R, for comparisons of medians. The Kendall rank correlation coefficient tau (τ), from the cor.test function in R, was used to quantify correlations. The dataset contained biological replicates from the same clinical participant. It was important to account for this non-independence of replicate samples in the final models. Generalised linear multi-level regression models were used for this purpose. They were constructed with bacterial load as the outcome, pregnancy outcome as the main predictor, and participant as a random intercept. Models, with and without this random intercept were compared to test whether such multi-level strategies were an improvement upon standard fixed effect approaches. The inclusion of participant number as a random intercept significantly improved the model fit. This is shown by a smaller Akaike Information Criterion (AIC) statistic for the multi-level versus fixed-effect model, which is statistically significant using a one-way ANOVA test (Table 5.1). Therefore, for all regression analyses, participant identity was included as a random intercept. Any additional covariates were included as fixed effects.

Table 5.1 – Comparison of fixed intercept versus multi-level regression model to account for variation in qPCR data.

Model	AIC	ANOVA P
Fixed intercept	1709.5	<0.0001
Random participant intercept	1685.5	

All regression models were checked for normality and homoscedasticity (equal variance) of residuals. Influential samples were identified as residuals with a Cook's Distance >0.1. Such samples were removed and the models rerun and any resulting change in effect sizes and/or significance recorded.

5.3 Results

5.3.1 Clinical dataset

Hospital records for all participating mothers and babies were available from the BBB database. Using these data, associations between our outcomes of interest and other clinical variables were investigated. Clinical characteristics that could confound the relationship between gestational age and bacterial load were identified via an extensive literature review, as documented in Chapter 1. Delivery method, maternal ethnicity, smoking, diabetes, and BMI were identified in this way and were extracted from the BBB clinical dataset. Associations between these potential confounders and preterm birth, or gestational age at birth, were then explored. Any such associations would inform downstream analysis and interpretation of experimental results. In order to improve power to detect such associations, all BBB clinical data that matched the cohort inclusion criteria were used in these preliminary analyses, regardless of whether a placental sample was available for experimental analysis.

The clinical and demographic characteristics of the clinical cohort and their relationships with pregnancy outcomes of interest are summarised in Table 5.2. The reliability of our clinical data was supported by the observation that the data reflected expectations. For example, boys in the dataset were significantly heavier than girls, and preterm infants had a significantly lower birthweight than those born at term. Sections 5.3.1-5.3.3 explore in further detail the nature of associations between GA at birth, preterm/term delivery, and *a-priori* identified confounders that were considered for inclusion in the final models.

Table 5.2 – Clinical and demographic characteristics of whole BBB preterm and healthy term cohorts.

Variable	Category	Sub-category	N (%)	Shapiro test (P)*	Median (IQR)	Mean	Range	Total (N)	Missing	Wilcox, X^2, or Fisher's exact test
GA at birth (weeks)	*All*			<2.2E-16	39 (38-40)	38.4	20-41	1961	0	
	Male			<2.2E-16	39 (38-40)	38.4	23-41	982	0	W=469140 P=0.50
	Female			<2.2E-16	39 (38-40)	38.4	20-41	939	0	
	Preterm			<2.2E-16	35 (32-26)	33.2	20-36	291	0	W=0 P<2.2E-16
	Term			<2.2E-16	39 (38-40)	39.3	37-41	1670	0	
Birthweight (grams)	*All*			<2.2E-16	3320 (2960-3650)	3230	295-5470	1923	38	
	Male			<2.2E-16	3010 (3010-3700)	3294	295-5470	977	5	W=410220 P=8.49E-5
	Female			<2.2E-16	3260 (2928-3600)	3175	340-4858	937	2	
	Preterm			0.0078	2167 (1566-2568)	2059	295-3910	282	9	W=25476 P<2.2E-16
	Term			7.58E-09	3400 (3120-3700)	3432	2008-5470	1641	29	
Maternal BMI (weight/ height^2)	*All*			<2.2E-16	24 (21-27)	24.77	14-66	1889	72	
	Preterm			3.05E-13	24 (21-27)	24.98	15-53	279	12	W=226790 P=0.79
	Term			<2.2E-16	24 (21-27)	24.73	14-66	1610	60	

Maternal age at booking	*All*			4.30E-09	33 (30-37)	32.8	14-48	1511	450	
	Preterm			0.046	33 (29-37)	32.7	17-48	257	34	W=1594 P=0.79
	Term			2.11E-08	33 (30-36)	32.9	14-46	1254	416	
Maternal diabetes	*All*	None	1417 (93.8)					1511	450	
		Essential	34 (2.3)							
		Gestational	60 (3.9)							
	Preterm	None	226 (89.0)					254	37	X^2= 12.1, P=0.0024
		Essential	10 (3.9)							
		Gestational	18 (7.1)							
	Term	None	1191 (94.8)					1257	413	
		Essential	24 (1.9)							
		Gestational	42 (3.3)							

Maternal parity	*All*	0	972 (50.8)		1914	47	
		1	648 (33.8)				
		2 or 3	266 (13.9)				
		4 or more	28 (1.5)				
	Preterm	0	151 (53.4)		283	8	Fisher's P= 0.19
		1	84 (29.7)				
		2 or 3	41 (14.5)				
		4 or more	7 (2.5)				
	Term	0	821 (50.3)		1631	39	
		1	564 (34.6)				
		2 or 3	225 (13.8)				
		4 or more	21 (1.3)				
Maternal smoking behaviour at booking	*All*	Non-smoker	1610 (88.2)		1825	136	
		Smoker	76 (4.2)				
		Quit <12 months	139 (7.6)				
	Preterm	Non-smoker	240 (87.0)		276	15	X^2=6.7 P=0.035
		Smoker	19 (6.9)				
		Quit <12 months	17 (6.1)				
	Term	Non-smoker	1370 (88.4)		1549	121	
		Smoker	57 (3.7)				
		Quit <12 months	122 (7.9)				

5.3.1.1 Maternal Ethnicity

Numerous studies have demonstrated a strong association between maternal ethnicity and sPTB incidence (Khalil et al., 2013, Moser et al., 2008, Muglia and Katz, 2010). There was a clear difference in the distribution of maternal ethnicities between the BBB preterm and term groups (Table 5.3). These observations are supported by previous data in which black women have higher rates of preterm birth than any other ethnic group (Muglia and Katz 2010). Asian women also had a relatively high rate of preterm birth in our cohort, which replicates independent data from a UK cohort (Moser et al. 2008).

Table 5.3 – Distribution of maternal ethnicities in whole cohort and preterm/term sub-groups. N (%).

	Asian, Asian British	Black, black British	White, white British	All others*	Missing	X^2
All	84 (4.6)	179 (9.9)	1226 (67.8)	318 (17.7)	154	
Preterm	19 (7.1)	47 (17.7)	146 (54.9)	54 (20.3)	25	X2=32.2 P=4.8E-7
Term	65 (4.2)	132 (8.6)	1080 (70.1)	264 (17.1)	129	

* Chinese, other Asian, other black, other, and all mixed groups

The estimated effect of maternal ethnicity on GA at birth in the cohort is presented in Table 5.4. Both black and Asian mothers had significantly shorter gestations than white mothers (Table 5.4). These groups also had significantly lower odds of having a term rather than preterm birth (Table 5.5). Similarly, Asian and black mothers had babies with significantly lower birthweights compared to white mothers (data not shown).

Table 5.4 – Linear regression for effect of maternal ethnicity on GA at birth

Predictor	β (95% CI)	β SE	P
Intercept	38.69 (38.54-38.85)	0.078	<2E-16
Maternal ethnicity (baseline = White, white British)			
Asian, Asian British	-1.37 (-1.29- -0.21)	0.31	1.03E-5
Black, black British	-1.26 (-1.24- -0.59)	0.22	1.03E-8
All others	-0.57 (-0.75- -0.068)	0.17	0.001

Table 5.5– Logistic regression for effect of maternal ethnicity on odds of having a term over preterm birth.

Predictor	OR (95% CI)	β SE	P
Intercept	7.40 (6.25-8.27)	0.089	<2E-16
Maternal Ethnicity (baseline = White, white British)			
Asian, Asian British	0.46 (0.27-0.81)	0.31	0.0051
Black, black British	0.38 (0.26-0.56)	0.22	4.19E-7
All others	0.66 (0.47-0.93)	0.17	0.017

5.3.1.2 Maternal BMI

Maternal BMI has been reported as a risk factor for PTB. There is a higher prevalence of PTB in both underweight and obese mothers (Girsen et al., 2016, Shaw et al., 2014, Torloni et al., 2009). To best model the non-linear relationship between GA at birth and maternal BMI, data were split into three categories. Underweight was defined as a BMI below 18.5 kg/m^2, normal weight as 18.5 to 19.9 kg/m^2, and obese as over 30 kg/m^2. In the BBB population there was no significant difference in the distribution of maternal BMI categories between preterm and term groups (Table 5.6). Lower GA at birth and reduced odds of having a term birth were observed for obese and underweight women, when linear and logistic models were run. However, these trends were neither statistically significant in unadjusted nor adjusted models (Appendix Table A 6 page 253 and Appendix Table A 7 page 253).

Table 5.6 – Distribution of maternal BMI categories in whole cohort and preterm/term sub-groups

BMI group N (%)	Underweight	Normal	Obese	X^2
All (N=1961)	57 (2.9)	1631 (83.2)	273 (13.9)	
Preterm (N=291)	11 (3.8)	232 (79.7)	48 (16.5)	X^2=3.01
Term (N=1670)	46 (2.7)	1399 (83.8)	225 (13.5)	P=0.22

5.3.1.3 Maternal diabetes

Maternal diabetes can increase the risk of preterm birth (Kock et al., 2010). Gestational or established maternal diabetes had a significantly reductive impact on GA at birth (gestational = -0.87 weeks, 95% CI= -1.62--0.12; established = -1.38 weeks, 95% CI = -2.36--0.39) (Appendix Table A 8 page 254). The odds of having a preterm rather than

term birth were also higher for women with gestational (OR = 0.44, 95% CI = 0.25-0.8) and established (0.46, 95% CI = 0.22-1.01) diabetes in this cohort (Appendix Table A 9 page 254). The inclusion of maternal ethnicity as a potential confounder in the above models, impacted effect sizes only marginally. However, the effect of gestational diabetes on GA at birth, and established diabetes on preterm versus term odds were no longer significant. Data on maternal diabetes status was relatively incomplete for our cohort. Its inclusion in downstream analyses would have significantly limited the sample size. Therefore, final models did not include diabetes, acknowledging that this may have limited the precision of the estimate. However, sensitivity analyses were run in which diabetes was included as a covariate in final models using the smaller sample set. Results were consistent whether or not maternal diabetes was included (data not shown).

5.3.1.4 Maternal smoking

Maternal smoking behaviour was associated with a significant reduction in GA at birth (Table 5.7), as well as birthweight (Table 5.8) in the cohort. Initial analyses were conducted using a three-group model for maternal smoking behaviour: women given up within 12 months of booking, smokers, and non-smokers. Smokers also had significantly higher odds of having a preterm birth than non-smokers (OR=0.53, 95% CI=0.31-0.92, P=0.019). However, no significant relationship was observed for those smokers who quit within 12 months, when compared to non-smokers.

Table 5.7 - Effect of maternal smoking behaviour on GA at birth using a linear regression model with categorical predictors to estimate effect size.

Predictor	β (95% CI)	β SE	P
Intercept	38.4 (38.27-38.54)	0.068	<2E-16
Maternal smoking (baseline=non-smoker)			
Smoker	-0.76 (-1.4--0.12)	0.32	0.019
Quit within last 12 months	0.13 (-0.35-0.61)	0.52	0.6

Table 5.8- Effect of maternal smoking behaviour on birthweight using a linear regression model with categorical predictors to estimate effect size.

Predictor	β (95% CI)	β SE	P
Intercept	3241.75 (3207.37-3276.13)	17.53	<2E-16
Maternal smoking (baseline=non-smoker)			
Smoker	-326.76 (-487.87--165.66)	82.14	7.23E-05
Quit within last 12 months	8.19 (-113.16-129.55)	61.88	0.9

GA at birth and birthweight were very strongly correlated in the cohort (τ=0.46, 95% CI=0.44-0.49, P<2e-16). The relationships between smoking and birthweight, and numerous other pregnancy complications, are widely reported (Mund et al., 2013). Therefore, further analyses were conducted in which the prediction of GA at birth from smoking status was adjusted for birthweight and vice versa. This analysis clarified that the main effect of smoking on fetal outcome in our cohort was on birthweight. The relationship between smoking and GA in our cohort may be confounded by the fact that lower weight infants tend to be born earlier. When birthweight is included in the model predicting GA at birth, the negative relationship disappears (β=0.25, 95% CI=-0.14-0.63) and is no longer significant. When the model predicting birthweight from smoking behaviour is adjusted for GA at birth, the effect of smoking is attenuated substantially, reducing the effect by almost 200 grams (β=-169.96, 95% CI=-267.66- -72.26). However, the association remains highly significant (P=0.00066).

A significant effect of smoking on birthweight is only observable for mothers who continued to smoke during pregnancy, not for those who gave up within the 12 months preceding booking. For this reason the two-group model of smokers and non-smokers was used in final models. This two-group effect on birthweight is clearly observed in Figure 5.2, which shows that smoking reduces birthweight, regardless of GA at birth. Previous evidence has described an association between smoking and GA at birth (Kyrklund-Blomberg and Cnattingius, 1998), as well as with microbiome profiles in sites such as the mouth (Wu et al., 2016). Therefore, although in the BBB data the majority of this association appears to be attributable to the confounding effects of birthweight, smoking was still considered an important covariate to retain in final models based on the strength of previous evidence.

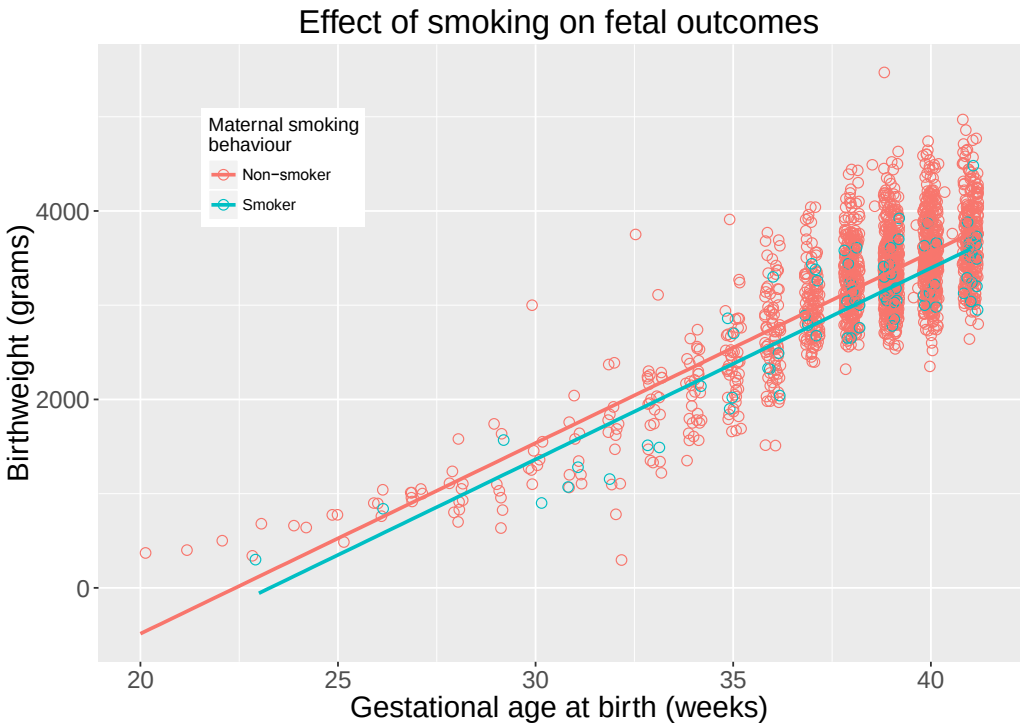


Figure 5.2 – Association between GA and birthweight by maternal smoking status. Lines indicate linear regression line for association between two variables for non-smokers (red) and smokers (blue).

5.3.2 Delivery method and gestational age at birth

Delivery method was identified as a potentially important confounder in associations between GA at birth and placental microbial profiles. In our data there was a trend towards more vaginal deliveries as GA at birth increased. However, the eight earliest births in this cohort were vaginal (Figure 5.3). This mild, but highly significant, positive relationship between increasing GA at birth and odds of vaginal delviery was confirmed using regression analysis (OR = 1.2, 95% CI = 1.15-1.25 per week increase). The association persisted and remained statistically significant with the addition of possible confounding covariates into the model (Appendix Table A 10 page 255).

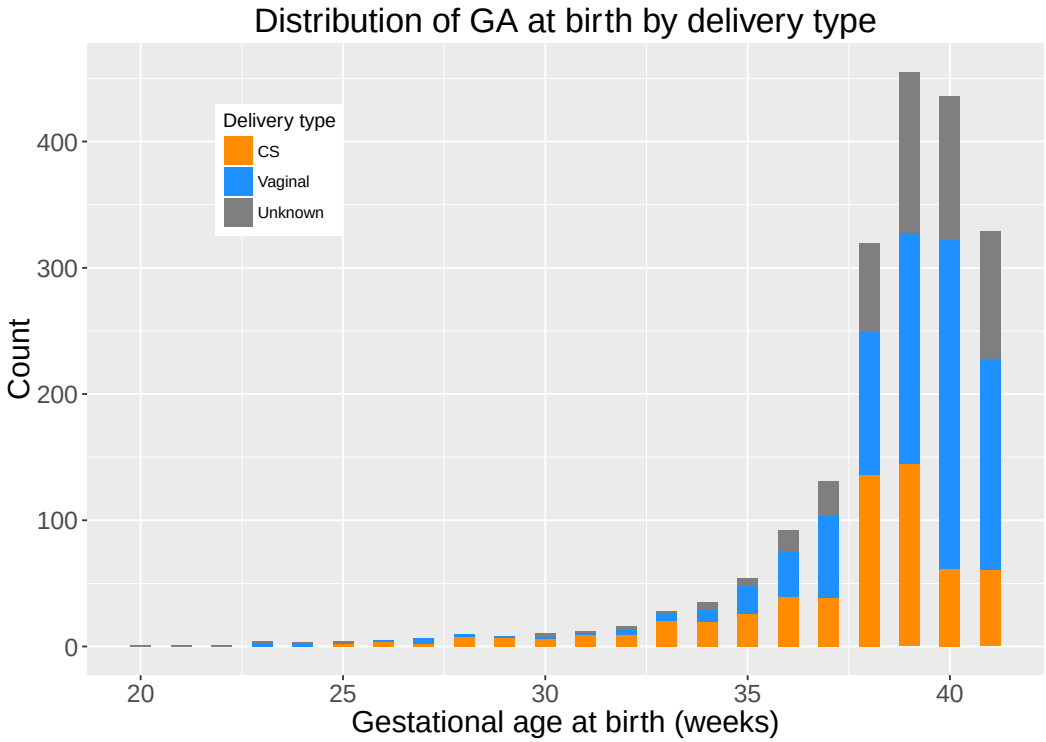


Figure 5.3 – Delivery method by GA at birth in the BBB preterm/healthy term cohort.

A significantly higher proportion of preterm births in the cohort were delivered by CS when compared to term deliveries (Table 5.9). This trend may reflect the higher incidence of clinical indications, such as PE, which require swift delivery to protect the health of mother and/or baby, in preterm rather than term groups. The odds of having a CS section and a preterm birth were 2.76 (95% CI = 2.09-3.66, P=1.07e-12) times higher than amongst term deliveries. This relationship was essentially unchanged by the addition of the potential confounding effects of maternal parity, age, BMI, and diabetes status into the model (data not shown).

Table 5.9 – Distribution (N (%)) of CS and vaginal deliveries in the whole preterm/healthy term BBB cohort and X^2 comparison of proportions.

Delivery method	CS	Vaginal	Missing	X^2
All	595 (40.2)	886 (59.8)	480	
Preterm	152 (60.8)	98 (39.2)	41	X^2=52.2 P=5.01E-13
Term	443 (36.0)	788 (64.0)	439	

5.3.3 Labour onset and membrane rupture

A central hypothesis of this thesis was that sPTBs would have a stronger bacterial signal than both nsPTB and term births. A sPTB is considered to be any delivery precipitated by spontaneous labour and/or spontaneous membrane rupture. Non-spontaneous events are those precipitated by artificial or no membrane rupture, combined with induced or no labour. These are considered inter-changeable with the 'indicated' PTB label. The proportion of non-spontaneous deliveries was significantly higher in the preterm than term group (Table 5.10), reflecting the fact that indicated preterm births make up the largest proportion of the BBB preterm cohort. Whilst a very small proportion of nsPTBs births occurred before 29 weeks (7%), 17% of sPTB events occurred in this very early period. This trend possibly reflects reluctance amongst obstetricians to facilitate delivery of very preterm infants, except in the most extreme cases. The very early births that do occur are more likely to be spontaneous, i.e. beyond medical control. Labour and membrane rupture data were not available for all BBB pregnancies. This limited the total sample size of certain final models that used sub-phenotype categorisations of preterm/term deliveries based on labour/membrane rupture data.

Table 5.10 – Number (%) of term and preterm births precipitated by spontaneous/non-spontaneous events in BBB preterm/healthy term cohort and X^2 comparison of proportions.

Outcome	Spontaneous	Non-spontaneous	Missing	X^2
Preterm	123 (48.6)	130 (51.4)	38	X^2=22.2 P=2.5E-6
Term	802 (35.4)	439 (64.6)	429	

Figure 5.4 and Figure 5.5 show the distribution of spontaneous/non-spontaneous labour and membrane rupture events by GA at birth in the clinical cohort. The figures are remarkably similar, reflecting the strong relationship between labour onset and membrane rupture method. The odds of having spontaneous labour occurring alongside spontaneous membrane rupture are 13.05 (95% CI = 10.23-16.74, P<2E-16) times higher than spontaneous labour associated with non-spontaneous membrane rupture.

Logistic regression analysis also revealed relationships between membrane rupture/labour onset and mode of delivery. The odds of having a vaginal birth following spontaneous rupture of membranes are 7.86 (95% CI: 6.21-10.01, P<2E-16) times higher than having a CS delivery following spontaneous rupture of membranes. In other words, spontaneous rupture of membranes (SROM) tended to be followed by a vaginal delivery. Similarly, the odds of having a vaginal birth following spontaneous labour were 9.52 (95% CI: 7.46-12.22, p<2E-16) times higher than having a CS birth following spontaneous labour. Whilst vaginal births may seed the infant and placenta with vaginal flora, CS deliveries are more likely to be seeded with skin flora. It is also plausible that differences in delivery method contamination may also have a quantitative component. These relationships with delivery method were critical to consider when analysing variation in placental bacterial load, according to definitions that used labour or membrane rupture data.

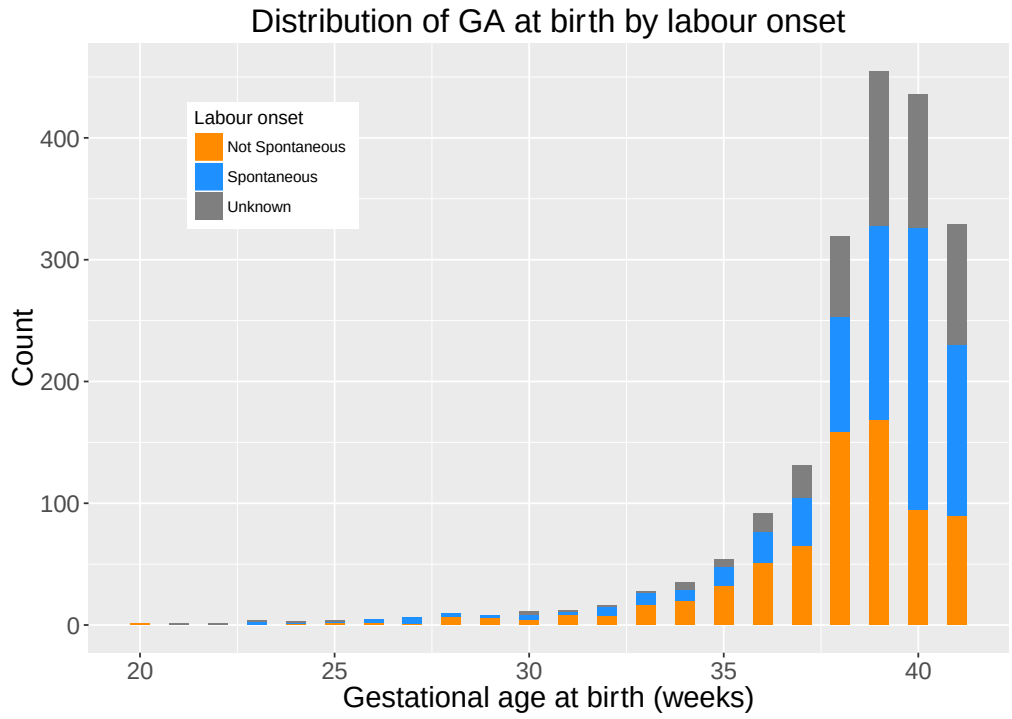


Figure 5.4 – Labour onset by GA at birth in the BBB preterm/healthy term cohort.

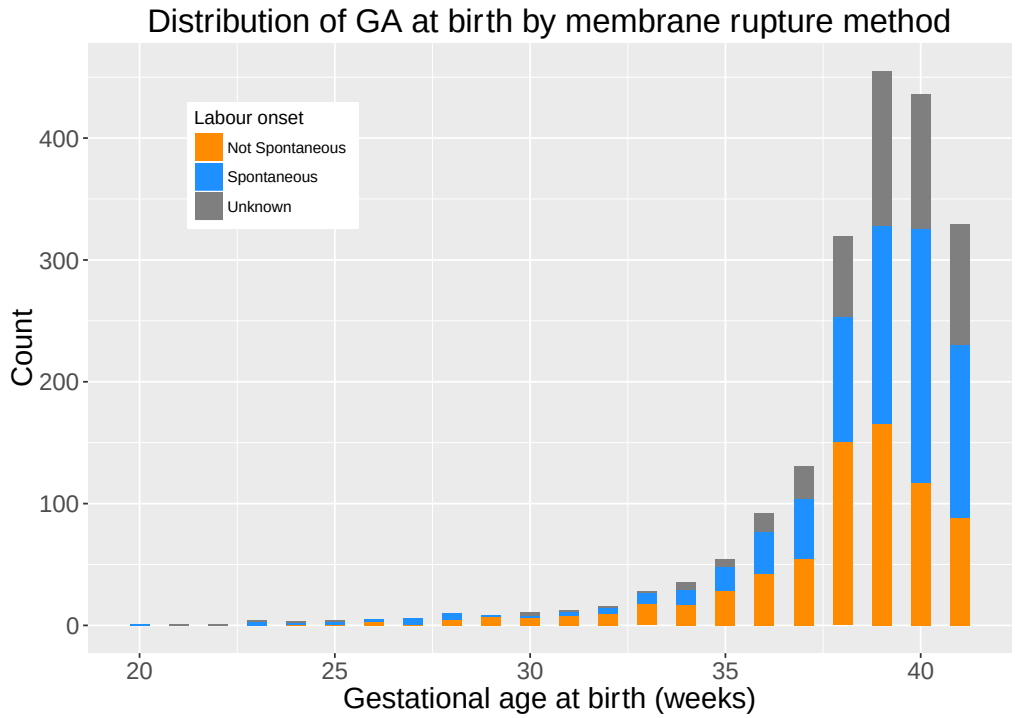


Figure 5.5 - Membrane rupture method by GA at birth in the BBB preterm/healthy term cohort.

5.3.4 BBB clinical data and sample availability

Placenta was not available for all BBB clinical data described above. It was important to establish whether any significant differences existed between the types of pregnancies for which placenta was, and was not, available. Delivery method, birthweight, GA at birth, labour onset, and membrane onset did not vary significantly between the BBB preterm samples with (N=146, 50.2%) and without (N=145, 49.8%) placental samples (see Appendix Table A 11, page 256). There were also no differences between these characteristics between term pregnancies with and without available placental samples. These observations imply that the reasons for which placentas were accrued were not biased by some clinically relevant outcome, such as delivery method or labour onset. All available preterm placenta were taken forward for qPCR analyses.

5.3.5 Summary of experimental cohort

This section will describe additional clinical data and sub-groupings for only those pregnancies within the experimental cohort, i.e. for samples with useable qPCR data. Preterm deliveries are routinely categorised into clinically relevant subgroups, for both clinical and research purposes. These divisions often include information regarding whether labour onset and/or membrane rupture method were spontaneous or not. In

this study, associations between pregnancy outcome and bacterial load were investigated using both categorical and continuous exposure variables. These were GA at birth; the binary category of preterm/term delivery; a three-group division of sPTB, nsPTB, and term delivery; and a four-group model of sPTB, nsPTB, spontaneous term, and non-spontaneous term outcomes. The distributions of spontaneous outcomes by preterm/term grouping in this qPCR cohort are outlined in Table 5.11. Further detail of the labour and membrane rupture processes underlying these groupings are in Table 5.12. It was also noted that 35 preterm membrane rupture events were prolonged (over 18 hours between rupture and delivery), and 8 term ruptures were prolonged.

Table 5.11– Distribution of spontaneous/non-spontaneous outcomes by preterm/term delivery in the experimental cohort (N (%)) and X^2 comparison of proportions.

Sub-group	Term (N=225)	Preterm (N=141)	X^2
Non spontaneous	73 (43.2)	70 (55.1)	X^2=3.7 P=0.056
Spontaneous	96 (56.8)	57 (44.9)	
Missing	56	14	-

Table 5.12– Detailed labour/membrane rupture groupings by preterm/term delivery in the experimental cohort (N (%)) and X^2 comparison of proportions.

Labour/membrane detailed	Term (N=225)	Preterm (N=141)	X^2
No labour/no membrane rupture	73 (43.2)	70 (55.1)	X^2=5.3 P=0.15
Spontaneous labour intact membranes	18 (10.7)	7 (5.5)	
Spontaneous labour/SROM	58 (34.3)	36 (28.4)	
SROM/no labour	20 (11.8)	14 (11.0)	
Missing	56	14	-

33 (23%) preterm infants in the experimental cohort were born under 33 weeks gestation, and 108 (77%) at between 33-36 weeks. 8 women who delivered preterm were given a cervical suture compared with 4 term deliveries. 23 preterm births were associated with preeclampsia (15.8%), 4 had congenital abnormalities, 2 resulted in a neonatal death, and 1 term birth was stillborn. 30 women who delivered preterm received steroids (20.6%). Table 5.13 summarises available clinical information on maternal infection in our experimental cohort from the clinical database. As can be seen, this information is relatively sparse and the vast majority of women had no recorded, clinically identifiable infection during pregnancy.

Table 5.13 – Maternal infections in experimental cohort by pregnancy outcome.

Maternal infection status	Term (N=225)	Preterm (N=141)
Asymptomatic bacteriuria	1	1
Candida	4	1
Chlamydia	1	0
Cystitis	1	0
Urinary tract infection	3	1
Group B Streptococcus	0	0
Other	1	2
None	161	120
Missing	53	16

Information regarding the level of neonatal care delivered to infants within the experimental cohort was available from the BBB clinical database. Over 50% of preterm infants, for whom data were available, ended up in special care (SCBU), or neonatal intensive care (NICU), units. By contrast, only 1.2% of term infants required this specialised attention (Table 5.14). 88% of infants with available data on their neonatal status who were born under 33 weeks were admitted to either NICU or SCBU, compared to 34% of 'late preterm' infants born between 33 and 37 weeks. These figures highlight the strong relationship between GA at birth and neonatal morbidity.

Table 5.14 – Level of neonatal care by pregnancy outcome (N (%)) and comparison of proportions using X^2 test.

Neonatal care	Term (N=225)	Preterm (N=141)	X^2
None	164 (98.8)	60 (48.8)	X^2=98.5
SCBU/NICU	2 (1.2)	63 (51.2)	P<2.2E-16
Missing	59	18	-

The distribution of samples collected at each of the three recruiting hospitals differed by pregnancy outcome (Table 5.15). Given the potential impact of variations in infection patterns, sterilisation procedures, and recruiter biases at different sites, on study outcome, recruiting hospital was included as a covariate in final models.

Table 5.15 – Distribution of preterm/term deliveries by study recruitment site (N (%)) and X^2 comparison of proportions.

Hospital	Term (N=225)	Preterm (N=141)	X^2
SMH	98 (43.6)	52 (36.9)	X^2=21.3 P=2.4E-5
CW	71 (31.5)	75 (53.2)	
QCCH	56 (24.9)	14 (9.9)	

5.3.6 Bacterial load and pregnancy outcome

Following exploration of the clinical dataset, summarised in preceding sections, experimental work to investigate associations between pregnancy outcome and placental bacterial load was carried out. Results from these tests are described and discussed in the remaining sections of this chapter.

Table 5.16 provides a summary of qPCR data across all samples, as well as by preterm/term groupings. As can be seen by the distributional characteristics of these data, and highly significant Shapiro Test output, all three were highly negatively skewed. Therefore, it was decided to log transform the 16S copy number data for downstream plotting, regression, and non-parametric comparisons. These summary data show distributions with large ranges that overlap between term and preterm groupings. Evidence for bacterial presence was observed across all types of pregnancies, even those with no pathological outcome. Indeed, the sample with the highest bacterial load, more than double the highest preterm value, was from a term delivered placenta with no recorded complication.

Table 5.16 – Summary of placental bacterial load data from qPCR experiments. Normality of distributions assessed using Shapiro test.

Group	Placental bacterial load (16S copy number)			
	Mean (SD)	Median (IQR)	Range	Shapiro test
All (N=701)	213.2 (714.0)	109.9 (70.8-179.9)	17.8-15860	W=0.15 P<2.2E-16
Preterm (N=265)	202.5 (508.1)	116.3 (71.8-192.6)	24.5-7348	W=0.22 P<2.2E-16
Term (N=436)	219.3 (814.5)	102.2 (70.8-167.7)	17.8-15860	W=0.13 P<2.2E-16

5.3.6.1 Bacterial load does not vary by mode of delivery

Placental load by delivery method was investigated in an attempt to explore whether bacterial load differed according to whether placenta were vaginally or abdominally delivered. Previous research has shown that vaginally delivered infants have a distinct structure of microbial colonisation compared to those delivered abdominally, via CS (Dominguez-Bello et al., 2010). However, whether this variation has a quantitative, as well as qualitative, component has not, to our knowledge, been explored. In our cohort, 605 experimental samples had data on delivery method available. No difference in

median bacterial load was observed between placentas delivered abdominally and those delivered vaginally (Figure 5.6). This was confirmed using both an unadjusted multilevel model, with participant number as a random intercept, and one in which batch and tissue type were also included as fixed level covariates (data not shown).

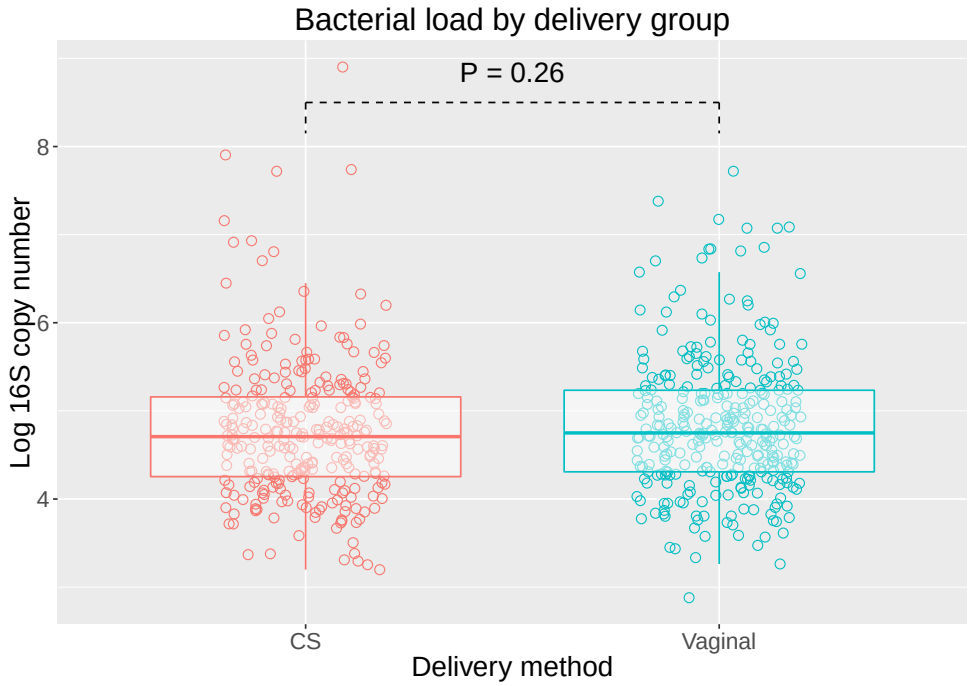


Figure 5.6 – Boxplot with individual points for log 16S copy number of placental DNA by delivery method. Boxes show median, 25th and 75th centiles with points beyond highest and lowest whiskers representing outliers (>1.5*IQR). P value for comparison of medians using Wilcoxon rank sum test.

5.3.6.2 Time between membrane rupture and delivery and bacterial load

Data for time between membrane rupture and delivery were available for 201 samples from 111 pregnancies. The maximum time between membrane rupture and delivery in the experimental cohort was 1875 hours (78 days). The mean was 87.6 hours (SD=273.2), and the median 16.6 hours (IQR 3.1-38.2). There was no significant association between time between rupture and delivery, and bacterial load, in unadjusted and adjusted models (data not shown).

5.3.6.3 Bacterial load and tissue type

Endogenous bacteria found at the maternal or fetal side of the placenta may have reached either site by distinct means. Organisms at the fetal side may have ascended

from the vagina and across the uterine tissues. By contrast, bacteria may be more likely to reach the basal plate from sites such as the oral cavity, via the maternal circulatory system. It is plausible that infection of the fetal side may be a more frequent event, reflected in a higher average bacterial load in these tissue samples. However, any such differences could be obscured in our dataset on account of delivery effects. Whilst parenchyma tissue was excised from beneath the placental membrane, the villous tissue was potentially more exposed to delivery contaminants because it was not covered in a protective membrane during delivery.

Running a univariate regression model, with participant number as a random effect, showed an increased load in parenchyma tissue, with an estimated reduction of -0.17 log 16S copy numbers in parenchyma samples (95% CI -0.33 - -0.016, P=0.031). However, when the same model was adjusted for batch and delivery method, this difference disappeared completely to 0.06 log (95% CI -0.1- 0.23, P=0.45). This shows that the univariate associations were confounded by delivery method and/or batch. Following adjustment, the null hypothesis of no association between bacterial load and placental tissue type could not be rejected.

5.3.6.4 Bacterial load does not vary by GA at birth or by preterm versus term outcome

The principal aim of this study was to investigate whether pregnancy outcome was associated with quantifiable differences in placental bacterial load. To address this aim, any associations with categorical and continuous predictors, based solely on gestational age data, were first investigated. Bacterial load did not significantly differ between preterm and term deliveries in the experimental cohort (Figure 5.7). Univariate and multivariate multi-level regression models were subsequently run to explore this association further. The same models were then run with the inclusion of only samples of parenchyma or villous tissue. There was also no association between bacterial load and GA at birth for adjusted and unadjusted models run on all samples or samples of either tissue type investigated individually (data not shown).

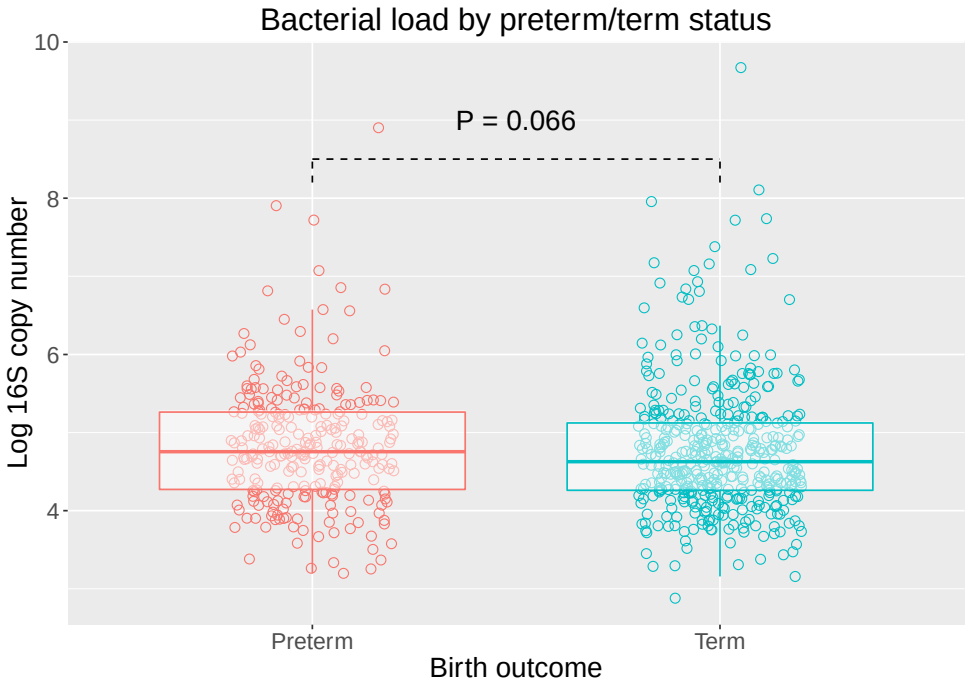


Figure 5.7 – Boxplot comparing log 16S copy number between preterm and term delivered placenta. Boxes show median, 25th and 75th centiles with points beyond highest and lowest whiskers representing outliers (>1.5*IQR). P value for comparison of medians using Wilcoxon rank sum test.

Previous research has reported an increase in incidence of infectious mediated PTB with decreasing GA at birth. Intra-amniotic colonisation is particularly frequent in the earliest births (<33 weeks) (Onderdonk et al., 2008a). In this cohort, there was no observable increase in bacterial load in the earliest births. These results were consistent whether a two group model, comparing early births to all other births, or a three group model comparing early preterm, to late preterm, and term births was used (data not shown).

5.3.6.5 Bacterial load is highest in placenta from spontaneous preterm births

An alternative approach to exploring associations between bacterial load and pregnancy outcome was to use predictor groups that took account of labour/membrane rupture information, in addition to GA data. Figure 5.8 summarises bacterial load using a four-group model of pregnancy outcome: sPTB, nsPTB, spontaneous term, and non-spontaneous term deliveries. Using a Wilcox-test for comparison of medians, it is clear that sPTB placental samples had a significantly higher median bacterial load, than each of the three other groups. This provides support for the hypothesis that placental 'infection', characterised by increased bacterial load, may have been involved in the aetiology of a proportion of sPTB events in this cohort.

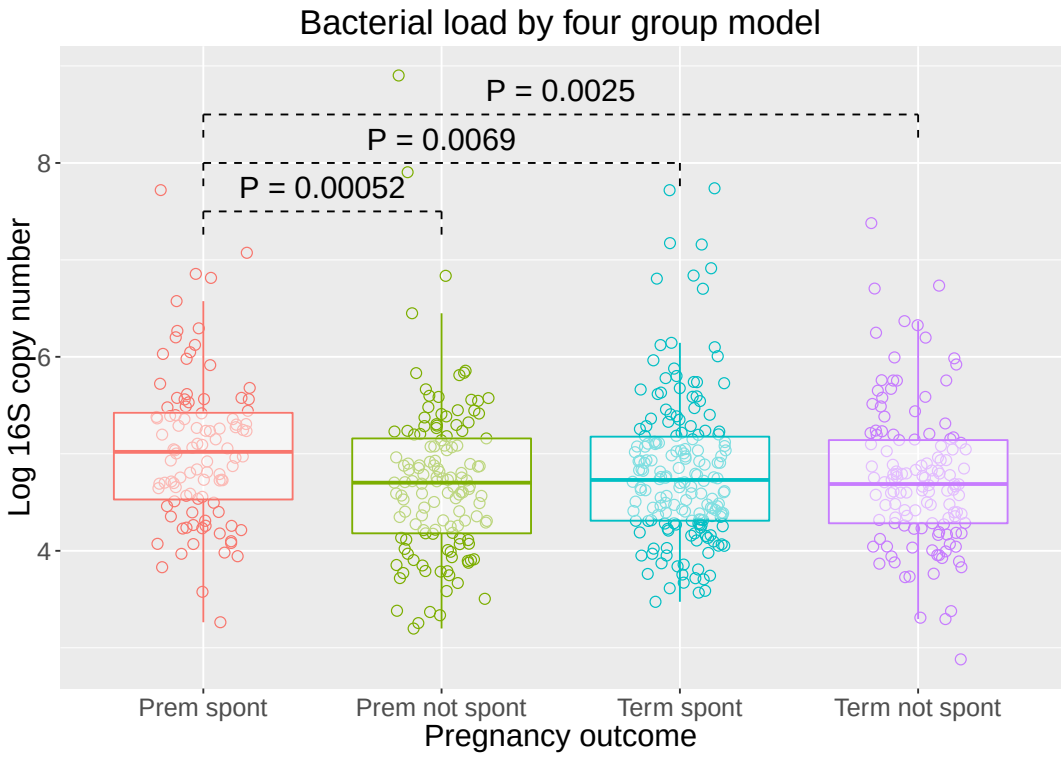


Figure 5.8 – Boxplot comparing log 16S bacterial load by pregnancy outcome (four group model). Boxes show median, 25th and 75th centiles with points beyond highest and lowest whiskers representing outliers (>1.5*IQR). P-value for comparison of medians using Wilcoxon rank sum test.

No difference was observed in bacterial load between the two term groups (W=11629, P=0.59). Multi-level regression analyses were conducted to further model any differences between the term outcomes and the null hypothesis of equal loads could again not be rejected. Therefore, it was decided to model associations between bacterial load and pregnancy outcome using a three-group, rather than four group model, in which both term groups were analysed together. When sPTB bacterial load was compared to data from all term births, power to detect the differences between the two groups was increased, as indicated by the smaller Wilcox Test P-value in Figure 5.9.

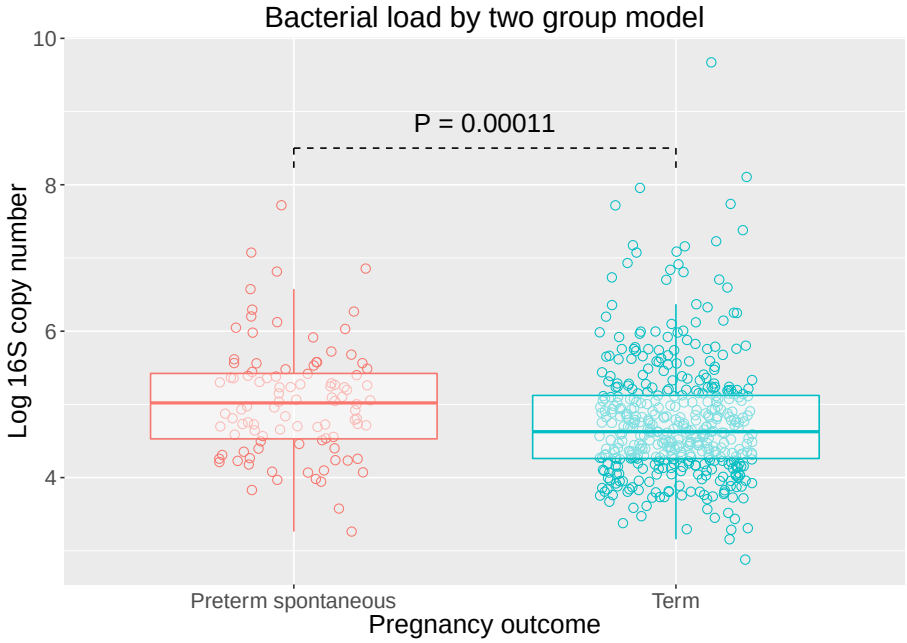


Figure 5.9 – Boxplots comparing bacterial load in placenta from sPTB and term pregnancies. Boxes show median, 25th and 75th centiles with points beyond highest and lowest whiskers representing outliers (>1.5*IQR). P-value for comparison of medians using Wilcoxon rank sum test.

Unadjusted regression models confirmed that placenta from sPTB deliveries had a higher bacterial load in comparison to nsPTB and term births. This trend was observed when all samples were analysed together in a single model. It was also observed when parenchyma, and villous tissue were analysed separately (Table 5.17). These differences were signfiicant for the whole cohort comparsions, and the parenchyma only model. When only villous samples were included, the difference was only signficant for the comparison between sPTB and term placenta.

Table 5.17 - Univariate models of association between bacterial load (log 16S copy number) and three group model of pregnancy outcome in whole cohort and parenchyma and villous tissue separately. Whole cohort and parenchyma analyses included participant identifier as a random intercept.

Predictors	All N=668				Parenchyma N=566				Villous=102			
	N	β (95% CI)	β SE	P	N	β (95% CI)	β SE	P	N	β (95% CI)	β SE	P
Intercept	-	5.03 (4.85-5.20)	0.089	<2.2e-16	-	5.04 (4.85-5.22)	0.095	<2.2e-16		5.02 (4.62--0.87)	0.20	<2e-16
Outcome (baseline= sPTB)	100	-	-	-	83	-	-	-	17	-	-	-
Preterm non spontaneous	133	-0.32 (-0.55--0.09)	0.12	0.0068	116	-0.34 (-0.58--0.09)	0.12	0.0071	17	-0.31 (-0.87--0.9)	0.28	0.27
Term	435	-0.26 (-0.45--0.06)	0.099	0.0095	367	-0.23 (-0.44--0.03)	0.11	0.027	68	-0.46 (-0.89- -0.02)	0.22	0.041

To investigate these associations further, the same models were run with the addition of fixed covariates: delivery method, maternal ethnicity, batch, maternal smoking, maternal BMI, and recruiting hospital (Table 5.18). The whole cohort model was also adjusted for tissue type. In the adjusted model for the whole cohort (N=569 samples, 308 participants), there was a significant reduction in placental bacterial load in nsPTB compared to sPTB placenta. The reduction in bacterial load in term births was attenuated in the adjusted model, and no longer significant. When two potential influential values from participants 2371 and 2081 were removed from the model, effect sizes were increased slightly, and significance for both comparisons increased. However, the comparison between sPTB and term load remained, marginally, non-significant (β nsPTB=-0.34, 95% CI -0.54--0.15, P=0.00058; β term = -0.15, 95% CI -0.31-0.0019 , P=0.053).

Similar effect sizes and patterns of signfiicance were observed in the adjusted models looking only at parenchyma tissue (N=488 samples, 299 participants) and those including samples from both tissue types. One influential sample from participant 2373, an nsPTB delivery, was identified via regression diagnostics. Its removal increased the strength of the associations for both groups (nsPTB = -0.36, 95% CI = -0.56--0.16, P=0.00042; term=-0.15, 95% CI = -0.32--0.0029, P=0.054).

The relationship between bacterial load and pregnancy outcome was investigated in villous tissue using an adjusted linear regression model. Again the trend was for sPTB placentas to have the highest bacterial load. However, in this relatively small sub-group analysis (N=81), these relationships were not significant. Removal of two potential influential values identified via diagnostic tests, lead to an increase in effect size and significance between nsPTB and sPTB samples (β =-0.72, 95% CI=-1.37--0.072), P=0.03). No change was observed with the comparison between sPTB and term placenta.

Table 5.18 – Summary of output from adjusted linear regression models predicting log 16S copy number from pregnancy outcome in full cohort, as well as parenchyma tissue, and villous tissue only. Full tables with data on all covariates included in models in Appendix Table A 12 and Table A 13, pages 257-258.

Analysis	Predictor	β (95% CI)	β SE	P
All samples	Intercept	4.67 (4.25-5.09)	0.21	<2.2e-16
	Pregnancy outcome (baseline=sPTB)			
(N=569)	**nsPTB**	**-0.29 (-0.5--0.08)**	**0.11**	**0.0064**
	Term	**-0.13 (-0.3-0.04)**	**0.09**	**0.13**
Parenchyma	Intercept	4.74 (4.32-5.16)	0.21	<2.2e-16
	Pregnancy outcome (baseline=sPTB)			
(N=488)	**nsPTB**	**-0.33 (-0.53--0.12)**	**0.11**	**0.0022**
	Term	**-0.15 (-0.32-0.02)**	**0.09**	**0.076**
Villous	Intercept	5.25 (3.72-6.78)	0.76	7.04E-09
	Pregnancy outcome (baseline=sPTB)			
(N=88)	**nsPTB**	**-0.59 (-1.43-0.26)**	**0.42**	**0.17**
	Term	**-0.32 (-1.01-0.36)**	**0.34**	**0.35**

5.3.6.6 Bacterial load and cervical suture

In some initial exploratory analyses an association was observed between whether or not a woman had a cervical suture during her pregnancy, and placental bacterial load (Figure 5.10). This was not a prior hypothesis. However, it was of interest to explore this further, given the clinical implications of any such association.

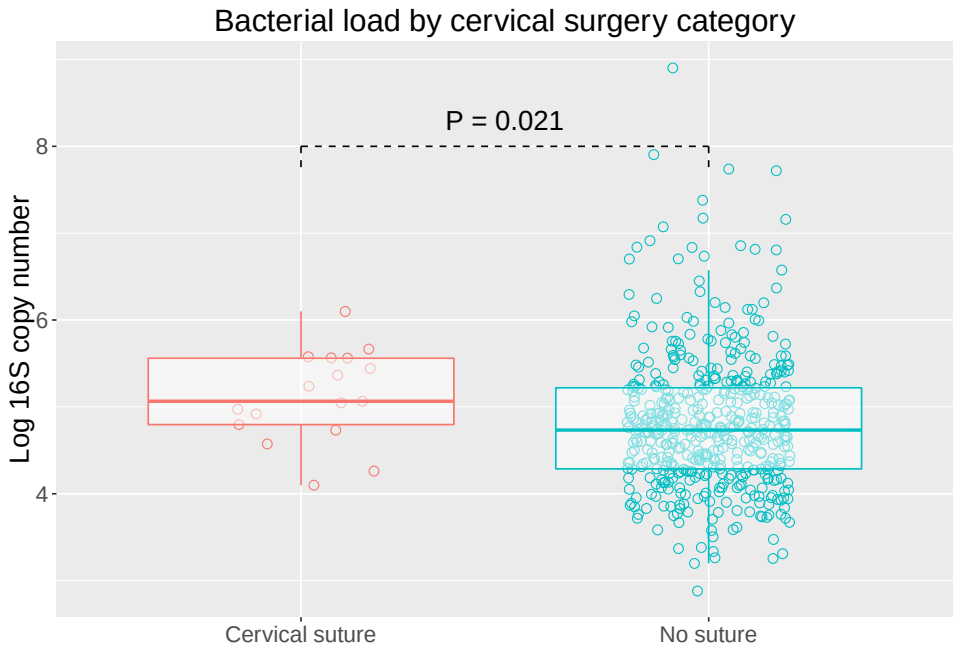


Figure 5.10 - Bacterial load appears to vary according to whether or not mothers have had a cervical suture in the current pregnancy. Boxes show median, 25th and 75th centiles with points beyond highest and lowest whiskers representing outliers (>1.5*IQR). P value for comparison of medians using Wilcoxon rank sum test.

This association was investigated further using unadjusted and adjusted multi-level linear regression models. 493 experimental samples from 273 pregnancies had associated information regarding whether or not a woman had a cervical suture during her pregnancy. Since only one of these samples was from villous tissue, it was decided to only conduct this analysis using the whole cohort, and not for each tissue individually. A reduction in bacterial load in placentas not associated with cervical sutures was observed using an unadjusted regression model, but this association was not significant (Table 5.19).

Table 5.19 – Unadjusted regression model for association between placental bacterial load (log 16S copy number) and cervical suture in current pregnancy.

Predictor	β (95% CI)	β SE	P
Intercept	5.12 (4.73-5.5)	0.20	<2.2e-16
Cervical suture (baseline = yes)			
No	-0.31 (-0.7-0.08)	0.20	0.12

The model was subsequently adjusted for the potential confounding effects of batch, delivery method, tissue type, and pregnancy outcome (three-group model). The effect remained the same and reduced marginally in significance, but the P value was still above 5% (Appendix Table A 14 page 259).

5.3.6.7 Bacterial load does not vary by level of neonatal care

Neonatal infection is dangerous, particularly for preterm infants, and such infections may originate from maternal uterine tissues or seeding at delivery (Camacho-Gonzalez et al., 2013, Chan et al., 2015). It was therefore important to investigate any associations between placental bacterial load and neonatal outcomes. Available clinical data on the level of neonatal care for each pregnancy was used as a proxy for outcome. 527 samples from 290 pregnancies were associated with information on neonatal care. No association was observed between whether or not a neonate received special care and bacterial load. This lack of association was confirmed using both unadjusted and adjusted multi-level regression models (data not shown).

5.3.6.8 Correlation between biological replicates is low

In this study, data on bacterial load from a substantial number of biological replicates, consisting of up to three samples from distinct sites on a single placenta, were available. By quantifying correlations between these replicates, the degree of similarity between bacterial profiles within individual pregnancies could be investigated. This provided a powerful opportunity to assess the extent to which placental 'infection' or overall microbiome load was a localised or diffuse phenomenon within placenta. Biological replicates could either both be from the fetal side of the placental disk (parenchyma 1 or parenchyma 2), or one from a parenchyma sample and the other from the maternal side of the placental disk (villous).

There was a significant and positive relationship between samples from the same participant, in which qPCR data from two parenchyma placental samples were available (Figure 5.11). However, this relationship was mild and appeared to be mainly driven by those samples with the highest bacterial load. To investigate this further, data was log transformed so that the most extreme values would have less influence on the overall correlation (Figure 5.12). Following transformation, the ICC remained significant, but reduced to 0.21 (0.08-0.33). However, after removing only two samples with the highest mean load between parenchyma replicates, this relationship disappeared, whether or not the data was log transformed (non-transformed ICC = 0.045, 95% CI = -0.084-0.17). This clearly indicates that the initial relationship observed between these replicates was solely driven by correlation between the most highly infected samples.

There was no observable relationship between bacterial load at the fetal side of the placenta (parenchyma tissue) and the maternal side of the placenta (villous tissue) (ICC=0.03, 95% CI=-0.17- 0.22). It was not possible to test correlations in bacterial load across the maternal side of the placenta because a maximum of one villous sample was ever run per participant.

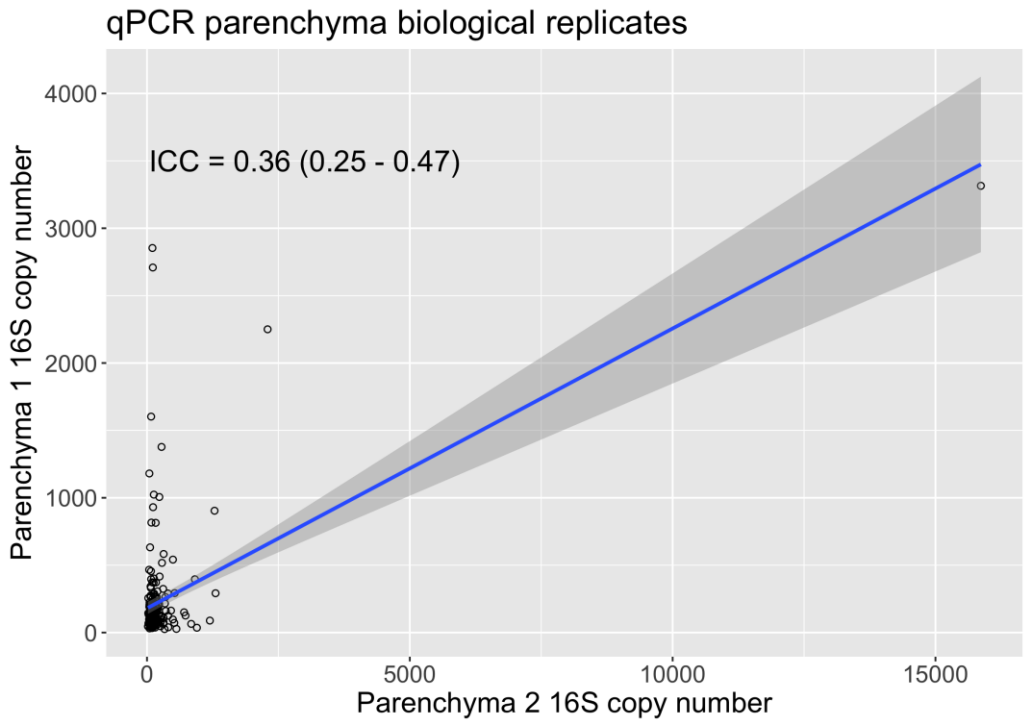


Figure 5.11– Relationship between bacterial load from biological replicates both taken from parenchyma samples from the fetal side of the placental disk.

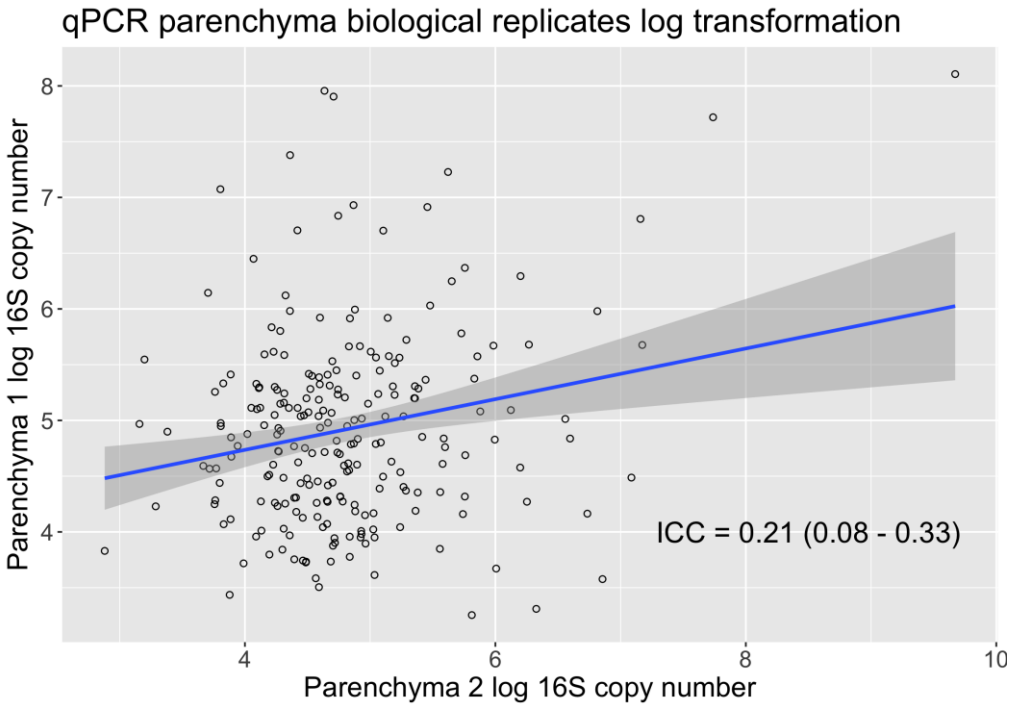


Figure 5.12 - Relationship between log transformed bacterial load from biological replicates both taken from parenchyma samples from the fetal side of the placental disk.

5.3.6.9 Negative extraction and PCR controls

Three PCR negative controls were run on each plate to control for the presence of contaminants in PCR reagents. In addition, a total of 21 'negative extraction' controls were run across the qPCR plates to explore the presence of bacteria in extraction kits. CT values from PCR negatives were consistently higher, indicating a lower load, than the experimental samples (W=11166, P<2.2E-16). In all but 8 of 699 total samples, sample CT values were over 1 CT cycle higher than the PCR negative controls on each respective plate. By contrast, there was substantial signal from the small number of extraction negatives. When compared with experimental samples, no significant difference in load was observed between the two (W=8622, P=0.17). This indicated the potentially significant impact of extraction procedures to bacterial signal in our study.

5.3.6.10 Summary of key results from qPCR analyses

A summary of the main findings discussed in this chapter is presented in Table 5.20.

Table 5.20 – Outline of key results from main statistical analyses conducted.

Association with placental bacterial load investigated	Significant difference using Wilcox test?	Significant trend in unadjusted regression model?	Significant trend in adjusted regression model?
Mode of delivery	No	No	No
Time between membrane rupture and delivery	No	No	No
Tissue type	Yes	Higher in parenchyma	No
Term/preterm birth	No	No	No
GA at birth	-	No	No
sPTB vs nsPTB /term	Yes	Higher in sPTB vs nsPTB	Higher in sPTB vs nsPTB
Cervical suture	Yes	No	No
Neonatal care	No	No	No

5.4 Discussion

In this study it was demonstrated that placental bacterial load was higher in sPTB deliveries, when compared to term and nsPTB events. This trend was observed when all samples were considered together, as well as separately by tissue type. Comparisons between sPTB and nsPTB events, using all samples and those from parenchyma tissue, were statistically significant in final adjusted models. No difference was observed in total load between placenta that were delivered vaginally and those delivered by CS, or when tissue from the maternal side was compared to tissue from the fetal side of the placenta. Preterm birth has previously been associated with increased frequency of intra-uterine infection using microbiological cultures, as well as broad-range and species specific end-point PCR (Jones et al., 2009, Onderdonk et al., 2008a, Han et al., 2006). This study is the first, to our knowledge, that demonstrates a quantitative association between total placental bacterial load and sPTB.

Cervical cerclage is a common prophylactic treatment for women at risk of preterm birth. However, literature on the associated risks of intra-uterine infection is sparse. Such risk could be due to either the morphological characteristics of the women that indicate them for suture in the first place, or the invasive nature of the procedure itself. A recent publication described dysbiosis in the vaginal microbiome among women who received braided over monofilament sutures, providing compelling evidence for the procedure's impact on maternal microbiome profiles (Kindinger et al., 2016). In the present study, there was an increase in median placental bacterial load associated with cerclage. However, potentially on account of the very small number of samples, regression models used to investigate this association were not significant. Replication of this comparison in a larger cohort is needed to confirm or refute this preliminary observation.

The fact that evidence for a positive association between bacterial load and sPTB was stronger when comparisons were made with nsPTBs, rather than with term births, is interesting. Arguably, nsPTBs are a better-matched control for the sPTB placenta, given that the placental samples are matched for GA at delivery. The non-spontaneous group could be seen as a representation of the infection status of 'normal' placenta at this stage of gestation. This provides support for the hypothesis that women who undergo sPTB events tend to be those exposed to a more profound bacterial insult prior to delivery. If the risk of ascending infection is a function of time, then *average* bacterial

load in the placenta may increase across gestation. This could explain the reduced difference between bacterial load in sPTB placenta and all term births, when compared to sPTB and nsPTB differences. However, such a positive association with GA was not observed in our experimental data.

From these data, as well as those from previous studies (Stout et al., 2013, Aagaard et al., 2014), it seems unlikely that all women who make it to term have sterile placenta. The similarity in signal between term and sPTB placental load may reflect a scenario in which those term delivered placenta with relatively high bacterial load were simply infected later in pregnancy. It is possible that a number of term births were precipitated by increased levels of bacterial placental colonisation that were causally involved in labour and/or membrane rupture. Because this challenge occurred later in gestation, it was not considered a pathological outcome. However, if this were the case we may expect to see an increased bacterial load in spontaneous term compared to non-spontaneous term events, which was not observed in this study.

A number of SNPs, particularly those in genes involved in immune related pathways have been associated with preterm birth (Crider et al., 2005). It may be that the divergent outcomes in women with similar levels of bacteria in their placenta could reflect important differences in maternal and/or fetal genetics. Certain women may respond to bacterial presence with an inflammatory signal that results in membrane rupture or labour, and preterm delivery. Others may be more likely to tolerate the foreign invasion longer and carry their infant to term.

In this study there was a measureable fluorescence signal, indicating the presence of bacterial DNA, across all placental samples. The interpretation of these signals will vary depending on the underlying assumptions of the nature of the 'placental microbiome'. If, as has been recently hypothesised by Aagaard et al. (2014), the presence of bacteria within the placenta is a normal and non-pathological occurrence, then a general bacterial presence across the whole cohort is to be expected, regardless of outcome. By contrast, it has traditionally been hypothesised that bacteria are only rarely able to break through the cervix and fetal membranes, eventually reaching the placenta. Under this scenario, it is more reasonable to consider only those samples with the highest loads to represent instances of 'true' bacterial infection. Signal from the remaining samples with lower loads, may instead be a function of stochastic noise and environmental/delivery contamination.

The placenta is an organ of very low biomass that, by definition, must pass through the bacterial rich environments of either the skin or the vagina prior to collection. The combination of these factors means that a large part of studying the 'true' placental microbiome must be developing tools to differentiate endogenous signal from organisms present in the organ during gestation, from noise picked up via reagent, delivery, and environmental contamination. A growing body of work is being developed to address these issues (Jervis-Bardy et al., 2015), with investigations concerning the lungs being of particular note (Aho et al., 2015). However, there is still relatively little literature directly addressing this topic.

In this study, bacterial load in negative extraction control samples was not significantly different to experimental samples. Clearly, this indicates that contaminating bacteria were present in extraction reagents, as has been noted in other publications (Salter et al., 2014, Weiss et al., 2014, Glassing et al., 2016). This observation also implies that much of the signal from the experimental samples may well be noise from contaminants picked up during sample preparation. However, it is important to consider, as discussed in section 3.2.2.1, the impact of background human DNA in this context. Competition with endogenous DNA was shown to affect the ability of the qPCR primers to pick up low biomass bacterial DNA in the mixed experimental samples. This makes it difficult to directly compare the quantitative results from negative extraction samples, in which no background contaminating DNA was present, with those in which the majority of nucleic acids present were non-target, human molecules, i.e. experimental placental samples. Therefore, despite the significance of the negative extraction qPCR signal, I propose that the comparisons within the experimental samples should be considered valid within the context of a potentially lower efficiency, mixed-sample qPCR assay. There also remains the likelihood that a number of samples in this dataset had very low, if any, true bacterial DNA present in-utero and that much of the signal is from organisms picked up at extraction or delivery.

It has also been shown in this study that bacterial load was only correlated between distinct samples excised from the same placenta in samples with the very highest load. Furthermore, no correlation was observed between biological replicates from opposite sides of the organ. These observations imply that where bacteria were present in the placenta, they were not necessarily diffuse across the whole organ, which is perhaps to be expected, except for in the most extreme cases. However, given the negative extraction observations, it is also possible that much of the signal in this assay was a

function of low level, stochastic noise originating from contaminants. If this were the case, this could also explain the low reproducibility between biological replicates.

Validating either hypothesis is challenging with only these quantitative data at hand and will be further investigated in Chapter 6. Using sequencing data, the qualitative nature of signals in all samples, both extraction negatives and placental extracts, will be explored. Part of the hypothesis for the proceeding work was that signal from extraction negatives would be of a different, non-clinically relevant make up when compared to those species identified in clinical samples of interest. Furthermore, although no difference in load between CS and vaginally delivered placentas were observed, qualitative differences by delivery group may be much clearer than quantitative ones. This is something that can only be explored through comparison of sequencing data.

There are several strengths to this study. Often studies investigating the molecular aetiology of preterm birth do not have such a large number of sPTBs to work with. Furthermore, given the completeness and comprehensiveness of the matched clinical data, important phenotypic distinctions could be investigated. It was also possible to include potential confounders in final models. The fact that negative extractions were investigated for their bacterial signal is also a strength of the study. This remains a rare practice within the field. Furthermore, the inclusion of technical replicates enabled the estimation of the reproducibility of the assay, which was shown in Section 3.2.2.2 to be high, with substantial correlation between such repeats. The nature of bacterial spread across the tissue was theoretically observable using biological replicates, although this was a more difficult analysis to interpret.

One of the main weaknesses of this study was that it used a discrete sample to make inferences about an entire organ. One, two, or three samples of less than 50 μg each were used to summarise the nature of bacterial colonisation across an organ that usually weighs about 500 gm. This relies on the assumption that bacterial presence will be relatively reproducible and diffuse across the organ, which is logically unlikely and not supported by our data. To overcome this limitation, taking a much larger number of samples, and mixing them together before DNA extraction, may enable a more reliable average estimate to be made. However, this then limits the chance to explore the nature of bacterial spread across the organ.

Chapter 6: The Placental Microbiome and Pregnancy Outcome

The principal objective of the work outlined in this chapter was to investigate whether the taxonomic bacterial composition of the placenta varied according to pregnancy outcome. It was shown in Chapter 5 that total bacterial load tended to be highest in placenta from sPTB placenta. The nature of this association was further explored using NGS techniques, and will be described in this chapter. Using 16S sequencing, the association between bacterial presence in the placenta and sPTB could be qualitatively explored. This served to elaborate in more detail, the quantitative associations reported in Chapter 5. The existence of a 'normal', shared placental microbiome, common across both healthy and complicated pregnancies, was also investigated. Dealing with issues of contamination in tissues of very low biomass, such as the placenta, were discussed in Chapter 3 and Chapter 5. These will be further considered in this chapter. Bacterial composition was addressed using a targeted 16S sequencing approach. The study cohort was a subset of samples from the quantitative characterization of placental bacteria and pregnancy outcome described in Chapter 5.

6.1 Introduction

As discussed in detail in Chapter 1, maternal infection is a well-known risk factor for sPTB. The investigations described in this chapter were carried out to explore this relationship, with respect to qualitative differences in bacterial infection by pregnancy outcome. A large challenge in the analysis and interpretation of these data was the differentiation of true, endogenous signal, from exogenous contamination and experimental artefacts. Table 6.1 outlines some of the potential sources of variation in 16S sequencing data. These alternative explanations must be kept in mind when analysing and interpreting results.

Table 6.1 – Potential endogenous and exogenous causes of observed variation in placental microbiome profiles assessed by 16S sequencing

Biological variation	Contamination	Experimental artefacts
Differences by ethnicity	Lysis and extraction kit contaminants	Primers chosen
Differences by outcome group	Delivery method	16S region targeted
Individual variation in microbiome profiles	Hospital specific contaminants	PCR error and bias
	Sample excision process	Sequencing error
	Storage buffer contaminants	Library quantification and normalisation techniques
		Chimera formation
		OTU picking strategies
		Reference database used for taxonomic classification

6.1.1 What we know about the placental microbiome

Evidence of bacteria in placental tissues comes from a number of sources: microbiological culture, PCR, molecular cloning and sequencing, fluorescence *in-situ* hybridization (FISH), histological staining, and most recently, next generation sequencing (Pelzer et al., 2016). These observations have often, but not exclusively, been associated with pathological outcomes, most commonly, PTB. The case for bacterial presence in the placenta is strengthened because evidence has come from both cellular and molecular techniques. The growth in the use of molecular techniques enables the identification of intra-uterine infections that often consist of predominantly anaerobic species that can be difficult to culture. By contrast, some studies have used cell based techniques to demonstrate the presence of live bacterial cells within placental tissue (Stout et al., 2013). These cell based techniques address a common criticism of purely molecular microbiome studies (Kliman, 2014). Specifically, that the presence of bacterial DNA does not prove that live, viable cells were present in tissue from which the DNA was extracted.

The majority of studies reporting bacterial presence in intra-uterine and placental tissue have been designed to investigate the aetiology of adverse pregnancy outcomes. However, bacteria have also been observed in uterine tissues from healthy pregnancies (Stout et al., 2013, Jones et al., 2009). Recently, Aagaard et al. (2014) described a

'unique placental microbiome' that was observed across both term and preterm pregnancies. These data have been used as evidence that a low biomass, non-pathogenic commensal community may be a functional component of normal human pregnancy. It has been suggested that such communities may train the fetal immune system prior to birth (Charbonneau et al., 2016). This has sparked considerable discussion in the literature regarding the meaning, reliability, and frequency of such non-pathogenic placental colonisation (Mysorekar and Cao, 2014, Wassenaar and Panigrahi, 2014, Payne and Bayatibojakhi, 2014, Charbonneau et al., 2016, Kliman, 2014). A recent study has challenged Aagaard et al.'s observations, using data from an independent, but small (N=12), cohort (Lauder et al., 2016). Lauder et al. (2016) observed no difference in the microbial signatures from contamination controls collected at all stages of collection and sampling, and DNA extracted from placental samples. These results suggest that previously reported signals did not originate from true placental colonisation, but instead from poorly controlled for contamination. It is plausible that observations of bacterial DNA within healthy placental tissue may reflect one of three scenarios:

1. An evolutionarily adaptation, facilitating beneficial exposure of the fetus to diverse microbes *in-utero,* during immune system development;
2. Neutral, stochastic, and low-biomass presence of organisms that have little or no functional significance and may never reach the fetus;
3. The result of a poorly controlled study in which most/all signal actually originates from exogenous contaminants.

The paper by Aagaard et al. (2014) described a community that consisted of low numbers of a variety of organisms. These organisms were largely from the phyla: *Bacteroidetes, Firmicutes, Fusobacteria*, and *Tenericutes*. The study also compared their data with those from the HMP (Huttenhower et al., 2012). This was an attempt to identify any overlap in composition with their placental cohort, and therefore identify the body site from which their sequences were likely to have originated. Their results were somewhat unexpected. They showed that placental samples clustered with oral, rather than vaginal samples from the HMP dataset. This has been used to support the claim that their observations are unlikely to reflect contaminants picked up at delivery,

but rather the hematogenous spread of microorganisms from the maternal oral cavity to the placenta (Aagaard et al., 2014). There has been previous evidence of the recovery of oral species from placental tissue (Han et al., 2010). However, these observations challenge traditional dogma in which the majority of organisms recovered from uterine tissues have tended to resemble the vaginal microbiome (Goldenberg et al., 2000).

The idea that the placenta may seed the developing fetus' environment and train its immune system is of great interest, whether or not the placental microbiome is established from the vaginal or oral communities. From where, and at what point in development, the fetus is first exposed to microorganisms, remains undetermined. First pass meconium has been shown to contain microbes, implying that there may exist some seeding prior to delivery in infants (Gosalbes et al., 2013, Hansen et al., 2015, Moles et al., 2013). Aagaard et al. (2014) use their data to suggest that a 'starter culture' of bacteria is transferred from mother to baby via the placenta *in-utero*. The significance of such a placental microbiome to fetal development is an active and nascent topic of research.

It is clear from studies examining infants delivered via the vagina or by CS that delivery method plays a large role in the primary seeding process of infants (Dominguez-Bello et al., 2010). This bacteria rich transition from the intra-uterine environment to the outside world also presented a methodological challenge in this study. Bacteria collected on tissues of interest during delivery may obscure signal from endogenous seeding that does occur. Experimental and computational methodologies for differentiating delivery contamination from endogenous signal require further attention.

The involvement of placental infection in adverse pregnancy outcomes, particularly preterm birth, is a less controversial topic. A number of microorganisms have been repeatedly implicated in the pathogenesis of preterm labour (Goldenberg et al., 2000). Many of these are of low virulence in normal situations, and are associated with BV. It is generally hypothesised that an opportunistic pathogen will ascend to the placenta from the vagina, and initiate preterm labour via inflammatory mechanisms (see Chapter 7). However, as mentioned above, the descent of organisms from the oral cavity to the uterus is an additional route that is gaining increasing attention. Different types of organisms may be found in the placenta, depending on whether they originate in the vagina or the oral cavity. Therefore, identifying whether vaginal or oral

colonisation routes are most likely to be responsible for intra-uterine infection, will improve the capacity of clinicians to target treatment more effectively.

The varying impact of different genera/species/serotypes upon obstetric outcomes has been noted in studies of both humans and animals (Oh et al., 2010, Migale et al., 2015). These varying responses highlight the importance of identifying which specific taxonomic groups are associated with adverse outcomes, rather than simply the identification of bacterial presence or absence. A detailed understanding of the types of microorganisms involved in sPTB will aid the development of more targeted clinical approaches. A summary of some of the key organisms that have been isolated from the intra-uterine environment, the majority using molecular techniques, and many of which have also been associated with sPTB, are listed in Table 6.2. This list is not exhaustive and is likely to grow as culture independent sequencing techniques continue to be employed. Given the potential influence of contaminants on low biomass tissues, such as the placenta, it is important to report associations with novel bacteria with caution.

Table 6.2 – Overview of most common organisms identified within uterine cavity tissues in normal and complicated pregnancies. AF=amniotic fluid; FM=fetal membrane; P=placenta

Genus	Example species	Tissue	References
Acinetobacter	*baumanii, lwooffii, junii*	AF and P	Prince et al. (2016); Romero et al. (2014d)
Atopobium	*vaginae*	AF, FM, and P	Jones et al. (2009); DiGiulio et al. (2010)
Bacteroides	*ureolyticus, fragilis, hemolyticus*	AF, FM, and P	Aagaard et al. (2014); Combs et al. (2014); Hitti et al. (1997);Han et al. (2009) Jones et al (2009); Marconi et al. (2011); Wang et al. (2013); DiGiulio et al. (2010)
Bergeyella	*zoohelecum*	AF, FM, and P	Aagaard et al. (2014); Combs et al. (2014); Han et al. (2009); Hitti et al. (1997); Jones et al. (2009) Marconi et al. (2011); Wang et al. (2013)
Clostridium	*hiranonis*	AF	Combs et al. (2014); Han et al. (2009); Wang et al. (2013); DiGiulio et al. (2010)
Enterococcus	*faecalis*	AF	DiGiulio et al. (2010); Markenson et al. (1997)
Escherichia	*coli*	AF, FM, and P	Hitti et al. (1997); Jones et al. (2009)
Fusobacterium	*nucleatum polymorphum*	AF, FM, and P	Aagaard et al. (2014); Combs et al. (2014); DiGiulio et al. (2008); DiGiulio et al. (2010); Gardella et al. (2004) Hitti et al. (1997); Han et al. (2009); Jalava et al. (1996); Jones et al. (2009); Romero et al. (2014c); Wang et al. (2013)
Gardnerella	*vaginalis*	AF, FM, and P	Combs et al. (2014); DiGiulio et al. (2008); Hitti et al. (1997); Jones et al. (2007); Romero et al. (2014c)
Haemophilus	*influenzae, haemoglobinophilus, parainfluenzae*	AF, FM, and P	Combs et al (2014); DiGiulio et al. (2010); Jalava et al. (1996); Jones et al. (2009); Wang et al. (2013)
Klebsiella	*pneumoniae*	AF and P	Aagaard et al. (2014); Hitti et al. (1997)
Lactobacillus	*crispatus, gasseri, delbrueckii*	AF, FM, and P	DiGiulio et al. (2010); Jones et al. (2009); Onderdonk et al. (2008b); Prince et al. (2016)

Leptotrichia	*amnionii*		Combs et al. (2014); DiGiulio et al. (2010); Gardella et al. (2004); Han et al. (2009); Marconi et al. (2011)
Listeria	*monocytogenes*	AF	Combs et al. (2014); DiGiulio et al. (2010); Wang et al. (2013)
Mycoplasma	*hominis*	AF, FM, and P	DiGiulio et al. (2008); DiGiulio et al. (2010); Hitti et al. (1997); Han et al. (2009); Jones et al. (2009); Marconi et al. (2011); Onderdonk et al. (2008b); Onderdonk et al. (2008a); Wang et al. (2013); Romero et al. (2014c)
Neisseria	*gonorrhoeae; cinerea; subflava; polysaccharea, lactamica*	AF and P	Aagaard et al. (2014); DiGiulio et al. (2008); DiGiulio et al. (2010)
Pantoea	*dispersa*	AF, FM, and P	Jones et al. (2009); Romero et al. (2014c)
Peptoniphilus	*lacrimalis, asaccharolyticus, harei*	AF, FM, and P	Jones et al. (2009); Wang et al. (2013)
Peptostreptococcus	*asaccharolyticus*	AF, FM, and P	DiGiulio et al. (2010); Hitti et al. (1997); Han et al. (2009); Jones et al. (2009)
Prevotella	*tannerae, oulora, bivia, copri, oris*	AF, FM, and P	Aagaard et al. (2014); Han et al. (2009); DiGiulio et al. (2008); DiGiulio et al. (2010); Hitti et al. (1997); Marconi et al. (2011); Onderdonk et al. (2008b); Wang et al. (2013)
Sneathia	*sanguinegens*	AF, FM, and P	Combs et al. (2014); DiGiulio et al. (2008); DiGiulio et al. (2010); Han et al. (2009); Romero et al. (2014c); Wang et al. (2013)
Staphylococcus	*equorum, pettenkoferi, aureus, hemolyticus*	AF and P	Combs et al. (2014); DiGiulio et al. (2010); Onderdonk et al. (2008b); Romero et al. (2014c)
Streptococcus	*agalactiae (Group B), thermophillus, oralis, mitis, anginosus, alivarius, pneumoniae*	AF, FM, and P	Combs et al. (2014); DiGiulio et al. (2008); DiGiulio et al. (2010); Han et al. (2009); Hitti et al. (1997); Jones et al. (2009); Prince et al. (2016); Romero et al. (2014c); Wang et al. (2013)
Ureaplasma	*urealyticum, parvum*	AF, FM, and P	Combs et al. (2014); DiGiulio et al. (2008); DiGiulio et al. (2010); Han et al. (2009); Jalava et al. (1996); Jones et al. (2009); Onderdonk et al. (2008b); Romero et al. (2014c); Wang et al. (2013)

6.1.2 The vaginal microbiome

The majority of intra-uterine infections are believed to be of vaginal origin. In order to colonise uterine tissues they must ascend from the lower genital tract, and break the functional and structural barrier of the cervix and mucous plug (Goldenberg et al., 2000). Therefore, understanding the nature of the vaginal microbiome before, during, and after pregnancy, is an important part of understanding infectious mediated PTB. The vaginal microbiome can vary by factors including age, ethnicity, menses, and sexual practices (Huang et al., 2014, MacIntyre et al., 2015). Additionally, changing environmental conditions, such as those associated with pregnancy, antibiotics, or hormone therapy, may facilitate normally commensal species becoming opportunistic pathogens. These changes may also enable the colonization of non-endogenous species within the vagina. Such changes could lead to situations that are detrimental to the host, such as reproductive tract infections.

A rich and complex ecosystem of microorganisms inhabits the vagina. Most members of this community are commensal species that are neutral to the host. Others are symbiotically beneficial mutualists, which are critical components of female reproductive health, most notably *Lactobacilli* spp. The dominance of *Lactobacilli* as the 'cornerstone of vaginal health' has been appreciated for over a century (Huang et al., 2014). This picture has been refined and developed using data from 16*S* and metagenome analyses of women, at various stages of their reproductive cycles and lives. One common framework describes five discrete and relatively stable vaginal 'community state types' (CSTs), which were first described in a group of asymptomatic, non-pregnant women (Ravel et al., 2011). Four of these CSTs were classified according to the dominance of a single *Lactobacillus* species: *L. crispatus* (CST I), *L. gasseri* (CST II), *L. iners* (CST III), and *L. jensenii* (CST V). CST IV is considered the somewhat anomalous state, and is often seen as a proxy for BV. It is characterized by a *Lactobacillus* poor community, and a relatively diverse, predominantly anaerobic, remaining microbiome. CST IV associated organisms include: *Prevotella* spp.*, Dialisters* spp.*, Atopobium vaginae, Gardnerella vaginalis, Megasphaera* spp.*, Peptoniphilus* spp.*, Sneathia* spp.*, Finegoldia* spp.*,* and *Mobiluncus* spp. (MacIntyre et al., 2015).

There is currently conflicting evidence regarding how variation in vaginal flora can impact pregnancy outcome (Romero et al., 2014b, Romero et al., 2014c, Hyman et al., 2014, Witkin et al., 2007). Studies into the relationship between BV and preterm birth

have formed a major part of this area of research. BV was traditionally considered to be the function of overgrowth of *Gardnerella vaginalis* and a concomitant reduction in the prevalence of 'healthy' *Lactobacilli* spp. However, recent evidence supports a definition of BV that is more polymicrobial, including organisms associated with CST IV and others such as *Ureaplasma* spp. (Africa et al., 2014). These relatively low virulence, mainly anaerobic organisms also tend to be the most commonly recovered species from intra-uterine samples. *Lactobacilli* spp., by contrast, are much less commonly recovered (Charbonneau et al., 2016).

Understanding why an environment rich in 'healthy' *Lactobacilli* spp. protects women from ascending infection could improve understanding of the link between infection and preterm birth. Figure 6.1 outlines a hypothesised mechanism behind *Lactobacilli* spp. dominance within the vagina (Zhou et al., 2007). During pregnancy, estrogen is produced by the placenta, which may further this dominance of *Lactobacilli* spp. and potentially encourage community stability. Indeed, the pregnant vaginal microbiome has been characterised as a period of relative stability, greater *Lactobacilli* dominance and reduced alpha, within sample, diversity (MacIntyre et al., 2015). Post-partum, estrogen levels fall dramatically, which may reflect some of the changes in microbiome dynamics observed in recent studies (MacIntyre et al., 2015).

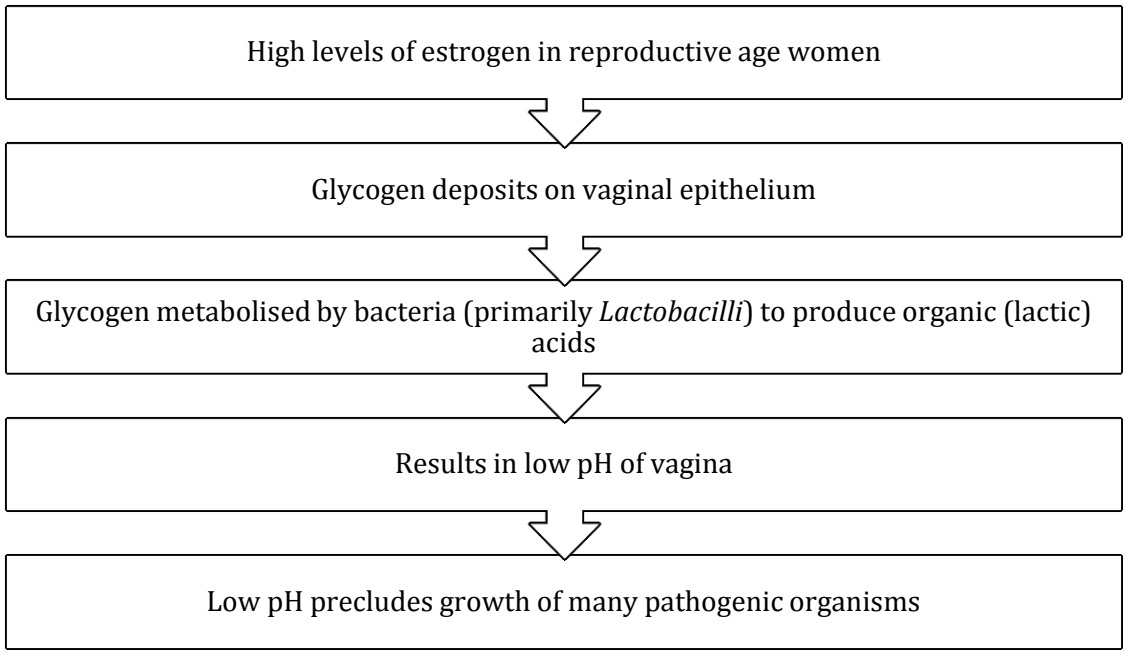


Figure 6.1 – Hypothesised link between estrogen, vaginal glycogen, *Lactobacilli* growth, and vaginal health (Zhou et al., 2007)

6.1.3 The oral microbiome and pregnancy

An association between maternal periodontal infection and PTB was first proposed two decades ago. Offenbacher et al. (1996) described a cohort in which periodontal disease was associated with low birthweight and PTB. Work that came out of this initial study indicated that this effect was likely due to hematogenous spread of oral microbes to the uterine cavity, independent of maternal vaginal and reproductive tract infection status (Offenbacher et al., 1996). An alternative explanation is that bacterial endotoxin or inflammatory cytokines originating at the oral site, rather than the organisms themselves, may have travelled to the uterine cavity and elicited an inflammatory response (Klebanoff and Searle, 2006). However, this explanation does not account for the genetic identity of specific bacteria recovered from both oral and placental sites in the same individual in more recent studies (Han et al., 2010).

Evidence from epidemiological studies, as well as animal models, has supported the role of periodontal disease as an independent risk factor for preterm birth (Offenbacher et al., 2006, Goepfert et al., 2004, Offenbacher et al., 2001, Arce et al., 2009, Boggess et al., 2005). The strength of association between periodontal infection and gestational length increases with disease severity, and reduces with gestational length. Crucially, these relationships persist following adjustment for other social risk factors that lead to poor oral hygiene and periodontal disease, such as smoking and access to care. Despite these associations, a relatively recent meta-analysis did not report an improvement in obstetric outcomes, following maternal periodontal disease treatment (Chambrone et al., 2011).

The most widely studied oral species implicated in preterm birth, as well as other adverse pregnancy outcomes such as stillbirth, is the gram-negative anaerobe *Fusobacterium nucleatum*. This organism is one of the most abundant commensal oral species, and is rarely detected at other body sites under normal conditions. A case-report provided the first direct evidence of translocation of the organism from a sub-gingival plaque infection to the placenta and fetus, eventually leading to inflammation related fetal death at term (Offenbacher et al., 2001, Jeffcoat et al., 2001). Its presence has been detected independently in a variety of uterine tissues including amniotic fluid (Wang et al., 2013), and placental and fetal membranes (Jones et al., 2009). It has also been found in the fetal lung and stomach (Han et al., 2010). The highly invasive nature of *F. nucleatum* has been investigated *in-vitro* and *in-vivo*. Such research has provided

functional support for its involvement in PTB, and its capacity to overcome the placental barrier between the maternal and fetal compartments. This work has also highlighted the involvement of the highly conserved protein FadA, a surface adhesin expressed by oral *Fusobacteria,* in intra-uterine infection. FadA is expressed on the organisms' cell surface. It has been shown to impact the ability of the bacteria to colonise and proliferate within the placenta in a mouse model (Han et al., 2010).

F. nucleatum infections are often part of a larger multi-species infection (Ikegami et al., 2009). It has been suggested that the binding of *F. nucleatum* to endothelial cells, via FadA, increases permeability of the endothelium to other organisms such as *E. coli* (Han et al., 2009). *F. nucleatum* may enable other microorganisms to spread in this way, facilitating mobile and mixed microbiological infections travelling to the intra-uterine cavity from distinct body sites. This is an interesting observation given that some estimates put poly-microbial invasion of the intra-uterine space at as high as 30% of cases (Fardini et al., 2011).

6.1.4 Hypotheses and objectives

This study was designed under the hypothesis that the placental microbial community would be altered in sPTB pregnancies when compared to term pregnancies. This alteration could have taken two main forms. Firstly, specific changes in the abundances of individual taxonomic groups, both at the OTU level, approximately accounting for species groupings, and the genus level, could predispose individuals to certain clinical outcomes. Linear models of normalized OTU count data were used to explore such associations between specific organisms and sPTB. Secondly, changes in the overall diversity of organisms within a sample could be associated with pregnancy outcome. For example, a more diverse environment may be more protective of adverse outcomes. In order to investigate this, alpha diversity estimates were used to assess whether pathological pregnancy outcomes associated with any changes in community diversity and richness. Beta diversity metrics were also used to visualize and quantify any potential differentiation in community structures between outcomes of interest.

The following questions address the main aims of this study:

1) Can we define a 'placental microbiome'?

2) Are there differences in the types of species recovered from vaginal and CS delivered placenta?

3) Are certain OTUs or genera enriched in sPTB, versus term and nsPTB delivered placenta?

4) Can we define a 'preterm placental microbiome' using ecological diversity metrics?

6.2 Materials and methods

6.2.1 Cohort selection

Time and resources permitted the sequencing of at least one sample from 73% of the pregnancies from the qPCR dataset. We plan to complete the sequencing dataset in the future when more resources become available. Figure 6.2 outlines the process of cohort selection and sample retention across all stages of my PhD, from the clinical analyses in Chapter 5, to beta diversity comparisons within this chapter. The proportions of samples sequenced from the three outcome groups were very similar between the qPCR and sequencing datasets (X^2=3.6, P=0.6), whilst fewer biological replicates were sequenced (Table 6.3). Given the smaller size of the sequencing cohort, both parenchyma (N=350) and villous (N=44) tissues were included together in final models, except for sections 6.3.4.2.2 and 6.3.4.2.3 where parenchyma samples were also analysed separately. An adjustment term was included in models to account for any potential confounding by tissue type, as well as a random intercept term to account for correlations between biological replicates. Alongside the experimental samples, 9 mocks, 19 negative extractions, and 5 PCR blanks were also sequenced. .

Table 6.3 – Distributions of samples by outcome group in sequencing and qPCR datasets.

Group	Sequencing dataset N (%)		qPCR dataset N (%)	
	Total pregnancies	Total samples	Total pregnancies	Total samples
sPTB	41 (16.5)	61 (15.5)	57 (15.6)	100 (14.3)
nsPTB	50 (20.2)	77 (19.5)	70 (19.2)	133 (19.0)
Term	145 (58.5)	237 (60.2)	225 (61.5)	435 (62.2)
Total	248*	394*	366*	699*

*Includes preterm samples with no delivery data

6.2.2 Sample collection and DNA extraction

See section 2.1.

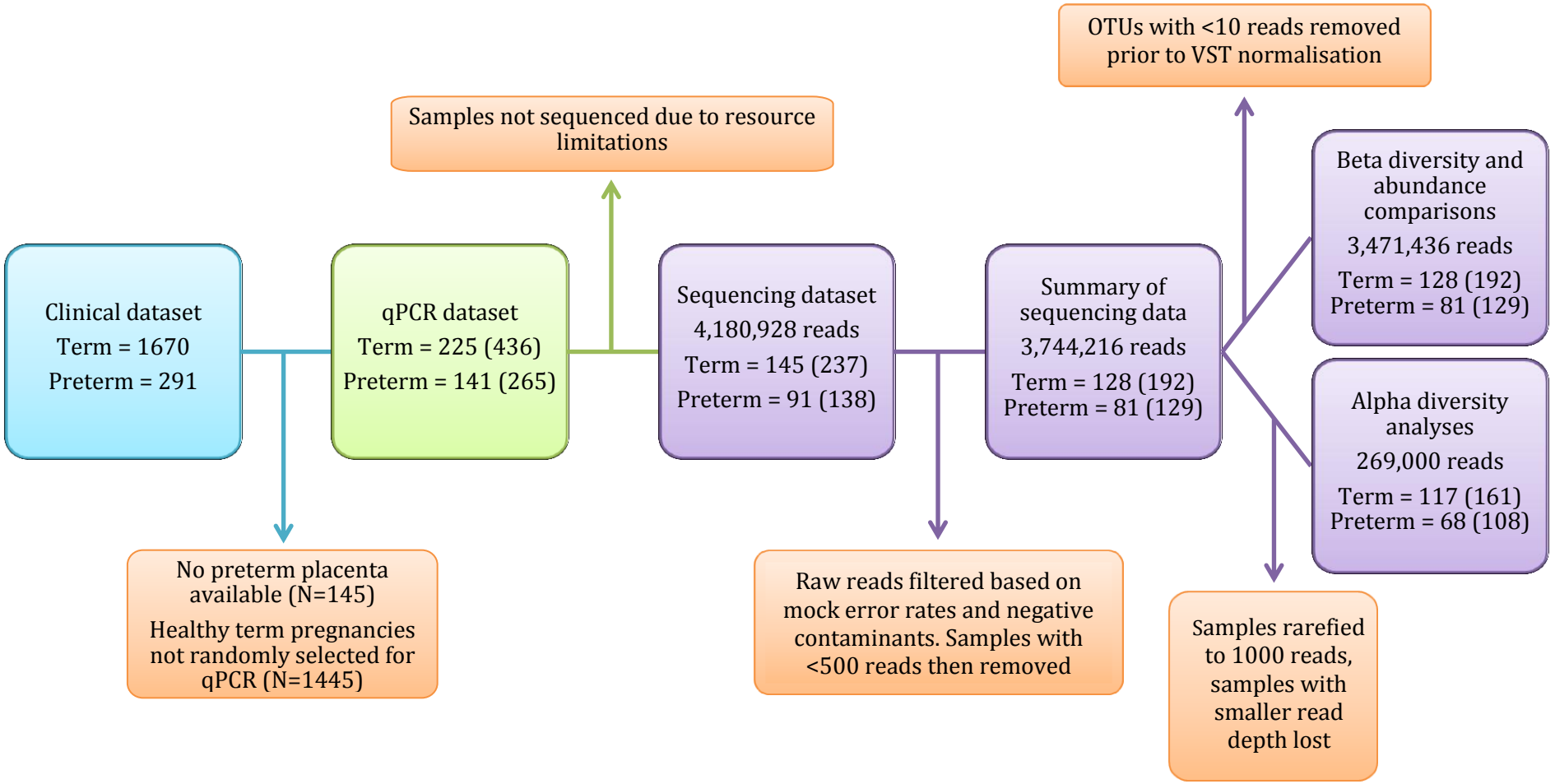


Figure 6.2 – Flow diagram documenting use and loss of samples (and sequencing reads) through process of PhD from clinical to final sequencing analyses. N total pregnancies (N total samples).

6.2.3 Quantitative polymerase chain reaction

See section 2.2.

6.2.4 16*S* rDNA amplicon high-throughput sequencing

See section 2.3.

6.2.5 Bioinformatic and statistical analyses

Following filtering, as described in sections 3.3.2.1-3.3.2.4, any sample with less than 500 reads was removed from the dataset. Taxonomic labels were assigned to the highest possible level using the QIIME pipeline. 64% of OTUs were assigned up to the level of genus, which accounted for 93% of filtered reads. Where OTUs of interest were identified, the representative sequence for this OTU was compared to the NCBI reference database, aiming to assign species level taxonomy to the group. Where a match of >97% to species in the NCBI database was observed, this was listed as the representative species for the OTU. Where there was ambiguity, a description only up until the level of genus was used. The Phyloseq v1.16.2 package in R was used to produce summary statistics, such as mean abundance by phyla or genera (McMurdie and Holmes, 2013).

6.2.5.1 Differential abundance testing: Limma

Differential abundance testing in microbiome datasets is often complicated by high variability in sampling sizes, i.e. read depths, and large numbers of zero counts. A number of different analytical strategies have been proposed to mitigate these issues. The majority of these build upon those developed for microarray and RNA-seq data. Most strategies involve some form of normalisation of count data to account for large variations in sampling depths, such as relative abundance or log transformation. There remains considerable debate regarding whether normalisation or rarefaction to even-sampling depth is the best strategy to deal with the statistical challenges of these datasets (see McMurdie and Holmes (2014) for discussion).

In this analysis, raw abundance data was normalised using the Variance Stabilizing Transformation (VST) approach in DESeq2 v1.12.4 (Love et al., 2014) within Phyloseq. These normalised data were then transformed into a Limma v3.28.21 (Ritchie et al.,

2015) object to take advantage of the multi-level functions available in this package. The package was first developed to model microarray data but has since been extended for use with other experimental designs. In order to improve the chances of identifying significant associations, and minimise the number of tests carried out, any OTUs unassigned at the level of genus and with 10 reads or less were removed. Models were corrected for the potential confounding influences of delivery method, maternal ethnicity, collection hospital, maternal BMI, tissue type, and maternal smoking. All models were run with a random intercept to account for correlations between biological replicates. P-values were corrected for multiple testing using the Benjamini-Hochberg procedure to produce Q values (Benjamini and Hochberg, 1995). OTUs were also merged to the level of genus and then models were re-run, to investigate genus, rather than OTU level, associations.

6.2.5.2 Calculating beta diversity

Three common methods for assessing distance or dissimilarity between samples or groups of samples are weighted UniFrac (Lozupone et al., 2007), unweighted UniFrac (Lozupone and Knight, 2005), and the Bray-Curtis dissimilarity metric (Bray and Curtis, 1957). Bray-Curtis is a quantitative, non-phylogenetic metric that falls between 0 and 1. It describes the degree of dissimilarity between the taxonomic structures of two communities, based on differences in absolute abundances. By contrast, UniFrac was a measure developed recently, which quantifies difference between two groups based on phylogenetic data. The measure counts the fraction of branch lengths in a phylogenetic tree leading to members of one group or other, but not to both. Weighted UniFrac also incorporates differences in relative abundances of taxonomic groups into the summary statistic.

Many approaches to using these methods suggest rarefying to an even sampling depth prior to calculating beta diversity. However, some have questioned the use of this technique given its requirement to remove data (McMurdie and Holmes, 2014). For beta diversity calculations, the VST normalised matrix was used. VST matrices can include negative values, which represent zero or very small original counts, and are not permitted by certain distance metrics. To mitigate this, any negative value was replaced with a zero, under the assumption that these cases were of very low, near zero abundance, and thus of negligible importance to the hypotheses under investigation. Following the computation of the three metrics, differences between the outcome

groups of interest were visually explored using Principal Coordinates Analysis (PCoA), a multidimensional scaling technique. The adonis function in the Vegan v2.4.1 package (Dixon, 2003) was used to quantify differences in beta diversity between outcomes of interest. Adonis uses a model analogous to multivariate analysis of variance (MANOVA), to assess how much variance in the beta diversity matrix can be explained by groups of interest. Significance tests are performed using *F*-tests, from 999 permutations of the raw data.

6.2.5.3 Calculating alpha diversity

Alpha diversity was calculated using the Shannon diversity index (Shannon, 1948), a summary statistic incorporating both overall diversity and evenness within samples. In order to account for any bias introduced by sequencing depth, whilst retaining absolute abundance counts, samples were rarefied to 1000 reads, over 100 iterations. The mean Richness and Shannon indices were then calculated. Associations between Shannon values, as well as raw richness calculations, with pregnancy outcomes were then investigated, using a mixed effects multivariate linear regression model, with participant identifier as a random intercept. All models were adjusted for the fixed effects of delivery, ethnicity, BMI, smoking behaviour, and tissue type.

6.3 Results

Following read and sample processing and filtering, 3,744,216 reads were retained, mapping to 298 unique biological samples from 215 pregnancies. A significant number of samples (N=96) were rejected during filtering, on account of low read depth (<500-fold). The mean read depth per sample following filtering was 11660-fold (sd=28137.4) and the median was 2977-fold (range=501-239500).

6.3.1 Read number and bacterial load

It was of interest to explore whether the sequencing and qPCR datasets were associated in any quantifiable way. When log bacterial load was used to predict log read depth in the cohort, there was a significant positive correlation between the two (Spearman's rank correlation coefficient=0.33, 95% CI=0.23-0.43, P=1.35E-9) (Figure 6.3). With each 1 log increase in read number there was a 0.5 log (95% CI=0.34-0.66) increase in 16S copy number (P=1.35E-9). This observation shows that there is a semi-quantitative component to sequencing output. It also suggests that those samples identified as most 'highly infected' by qPCR data, had the most amount of sequencing information retrieved from them.

6.3.2 Summarising the 'placental microbiome'

99.98% of reads matched to a taxonomic classification at the level of Phyla. The most abundant phylum recovered from all placental tissues was the *Firmicutes*, accounting for 70% of filtered reads. The distributions of the top 5 phyla recovered from placental tissue, accounting for 99% of reads, are described in Table 6.4. Trends in total load did not always mirror those relating to average abundance. For example, the *Tenericutes* had the third largest number of reads overall, but the 5th largest mean relative abundance. Whilst some samples were almost entirely made up of *Firmicutes, Proteobacteria* and *Tenericutes,* by contrast*, Bacteroidetes* and *Actinobacteria* never made up more than 38 or 76% of any one sample, respectively. Variation in individual placental sample composition is demonstrated using examples from four pregnancies in Figure 6.4. These plots also demonstrate the similarity in phylum level compositions across placental biological replicates taken from the same pregnancy.

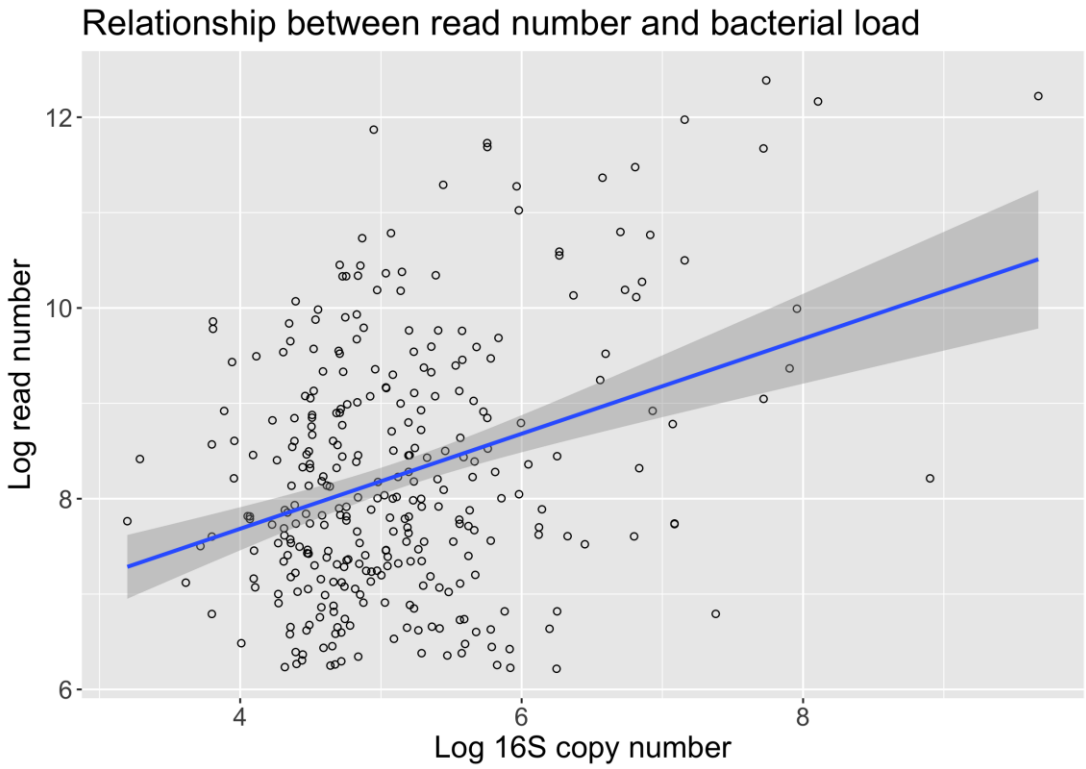


Figure 6.3 – Bacterial load is positively associated with the total read number from 16S sequencing analysis. Individual samples with regression line and 95% confidence interval around the estimate are shown.

Table 6.4 – Total read and relative abundance distributions across the top 5 phyla in order of total abundance.

Phylum	Total reads (% total)	Mean (SD) total reads	Median (range) total reads	Mean (SD) % relative abundance	Median (range) % relative abundance
Firmicutes	2631043 (70.3)	8196 (26741)	1072 (2-239400)	46.3 (28.3)	42.4 (0.035-99.96)
Proteobacteria	500476 (13.4)	1559 (4354)	576 (1-46670)	29.4 (25.0)	24.3 (0.0009-99.46)
Tenericutes	249100 (6.7)	776 (6698)	0 (0-86120)	3.0 (12.8)	0 (0-99.17)
Actinobacteria	245924 (6.6)	766 (1596)	335 (4-15840)	16.3 (13.9)	13.5 (0.0058-76.04)
Bacteroidetes	79814 (2.1)	249 (1041)	33 (0-13770)	3.3 (6.0)	0.74 (0-37.37)

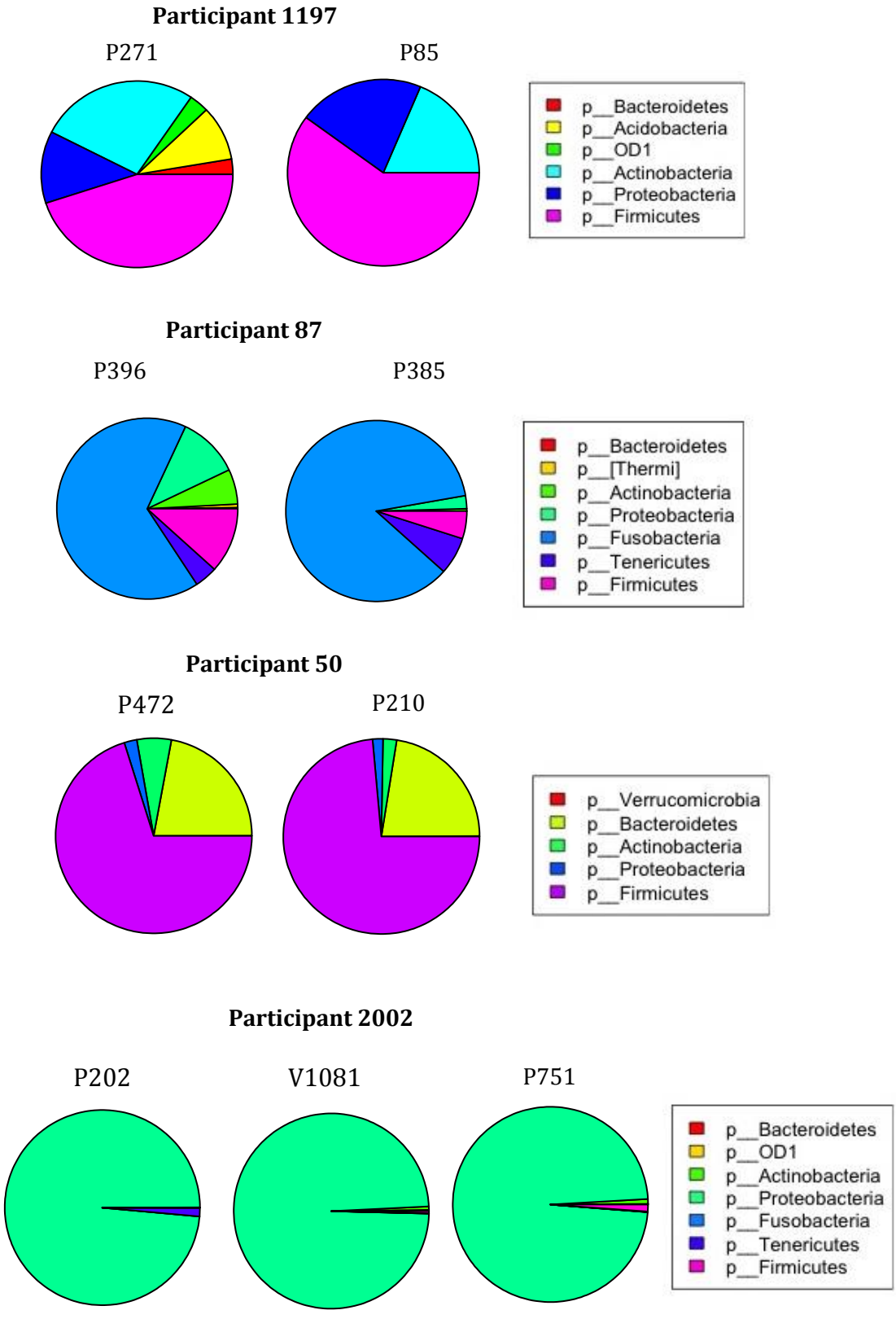


Figure 6.4 – Phylum level composition of individual placental biological replicate samples, taken from four different participants. P = parenchyma, V= villous.

92.8% of reads mapped to a genus using the QIIME pipeline. The distribution of the read counts for the top 10 genera were slightly different when ordered by total reads (Table 6.5) or mean relative abundance (Table 6.6). Most notably, *Veillonella* is the third most abundant genus in terms of total reads, but its mean relative abundance is only 1.02%, 12th in total. Similarly, *Mycoplasma* does not appear in Table 6.6, although it is the fifth most abundant genus in Table 6.5. The genera in Table 6.5 account for 78.4% of total reads in the dataset. Whilst the first three genera In Table 6.5 (*Streptococcus, Lactobacillus*, and *Veillonella*) represent almost 60% of total reads, the remaining genera in this dataset were of relatively low total abundance. The median number of reads in any genus was 284, but the highly skewed nature of the dataset meant the mean was significantly larger at 9001. This number was further reduced when individual OTUs were considered. The median number of reads in a single OTU, post filtering, was just 34, and the mean was 1266. Theoretically, OTUs in this dataset should approximate species level assignments, given the 97% similarity threshold used in OTU clustering. However, it is likely that phenomena such as sequencing artefacts and PCR error, lead to some inaccurate and redundant divisions in this dataset. Indeed, when comparisons of interest were made to the NCBI database, many unique OTUs mapped to the same species classifications.

Another way of describing the taxonomic make-up of the placental microbiome in this cohort was to use rank abundance. The top 20 OTUs in Figure 6.5 were all present in over 25% of samples, demonstrating that a substantial number of OTUs were shared across a substantial proportion of samples. OTU 1098473, mapping to *Propionibacterium acnes* was present in 75% of samples. The observation of many shared OTUs across the dataset could be interpreted as providing support for the existence of a shared, unique placental microbiome as suggested by previous groups (Aagaard et al., 2014). However, it is worth noting that many of these shared OTUs map to taxa that are skin and vaginal commensals. This raises the alternative possibility that this signature may largely reflect delivery contamination, rather than the actual endogenous placental microbiome.

Table 6.5 – Top 10 most abundant genera in order of total abundance.

Genus	Total reads (% total)	Mean (SD) reads	Median (range) reads
Streptococcus	1074952 (28.7)	3349 (19433.3)	86 (0-239000)
Lactobacillus	762789 (20.4)	2376 (10915.6)	35 (0-122900)
Veillonella	393271 (10.5)	1225 (15340.5)	0 (0-200800)
Staphylococcus	155937 (4.2)	486 (1539.6)	129 (0-20290)
Ureaplasma	152502 (4.1)	475 (4943.2)	0 (0-86120)
Mycoplasma	95334 (2.5)	298 (4499.9)	0 (0-79190)
Erwinia	88754 (2.4)	277 (2804.3)	0 (0-34350)
Corynebacterium	79269 (2.1)	247 (529.7)	100 (0-4615)
Pseudomonas	71448 (1.9)	226 (2170.4)	3 (0-37870)
Propionibacterium	60537 (1.6)	189 (483.1)	65 (0-5443)

Table 6.6 – Top 10 genera in order of mean relative abundance

Genus	Mean % relative abundance (SD)	Median % relative abundance (range)
Lactobacillus	16.15 (27.2)	1.1 (0-99.2)
Streptococcus	11.27 (20.9)	3.1 (0-99.8)
Staphylococcus	10.53 (15.0)	4.9 (0-79.7)
Corynebacterium	6.75 (7.8)	3.7 (0-43.2)
Propionibacterium	5.33 (7.2)	2.7 (0-60.6)
Escherichia	4.9 (14.0)	0 (0-99.1)
Acinetobacter	4.2 (11.3)	0.7 (0-96.5)
Sphingomonas	4.1 (9.9)	0 (0-59.0)
Pseudomonas	3.54 (10.6)	0.04 (0-98.7)
Ureaplasma	2.56 (11.7)	0 (0-89.7)

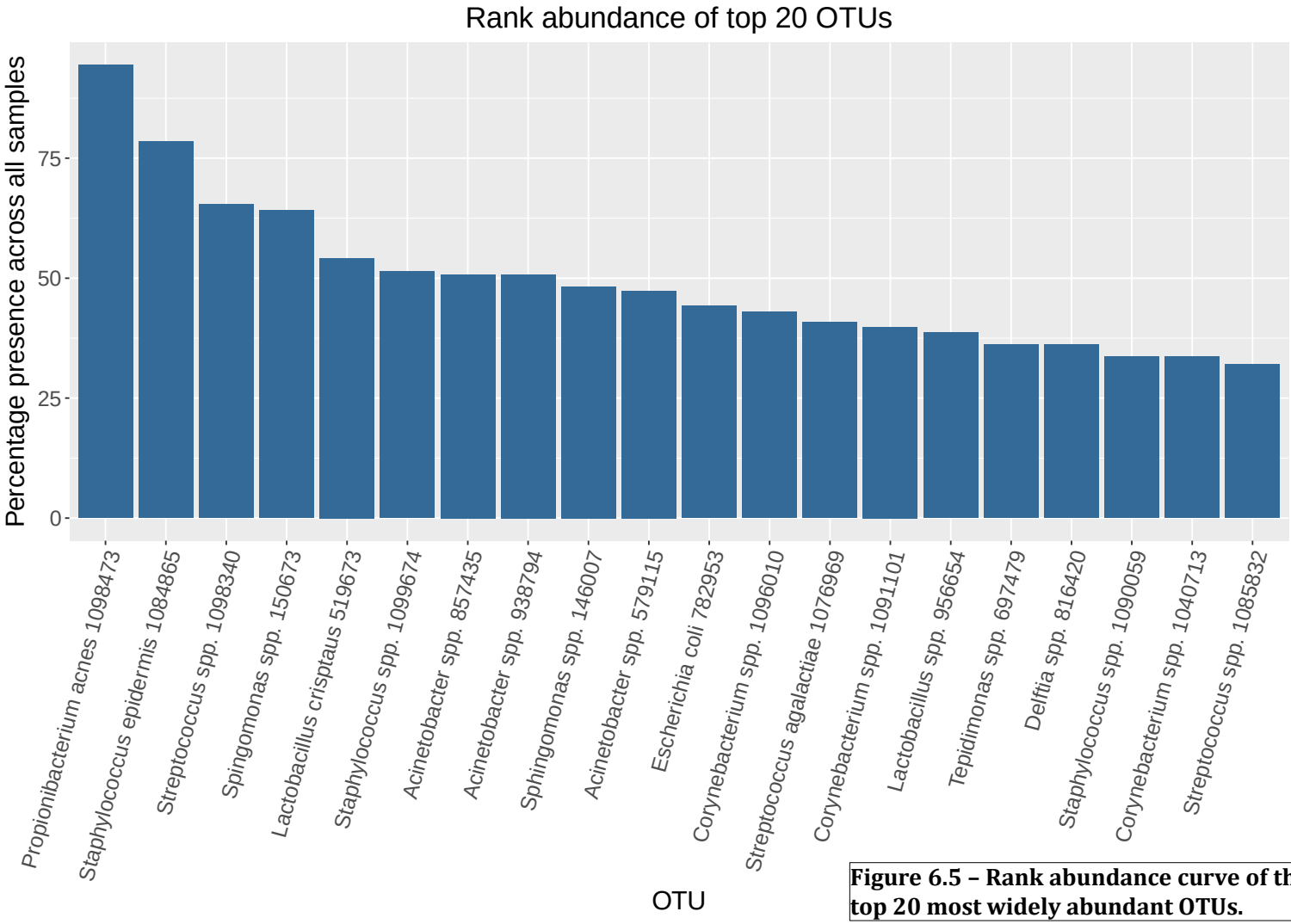


Figure 6.5 – Rank abundance curve of the top 20 most widely abundant OTUs.

6.3.3 Delivery method is very influential for certain key genera

Around half of the placental samples in the cohort contained reads mapping to the vaginal commensal species *Lactobacillus crispatus* (Figure 6.3). Given that 47% of deliveries in this cohort were vaginal, this observation warranted further exploration. In order to investigate the potentially significant contribution of delivery contaminants to sequencing results, the taxonomic make-up of vaginal and CS deliveries were compared graphically and statistically.

Figure 6.6 combines the ten genera with highest mean relative abundance from CS deliveries with the ten highest from vaginal deliveries. Most of these genera were shared between CS and vaginal groups. However, when the mean relative abundances were compared between the two groups, there were some significant differences, most notably in the *Lactobacilli* distribution. Whilst in vaginal samples, the mean relative abundance of *Lactobacilli* spp. was over 0.25, in CS delivered placenta it was under 0.05. When the relative abundances of all these 11 genera are compared between the two groups, using an independent two-sided T-test, 6 are significantly differentially abundant (Table 6.7). Two common vaginal genera, *Ureaplasma* and *Lactobacilli* were significantly more abundant in vaginal deliveries. By contrast, the common skin flora, *Streptococci, Staphylococcus, Propionibacterium*, and *Corynecbacterium,* were present at significantly higher abundance in CS samples. These observations confirm that certain highly abundant genera likely reflect delivery method, rather than endogenous colonisation. Controlling for delivery method is therefore critical in order to tease out any potential signals from organisms that may be associated with pregnancy outcome. For example, *Ureaplasma* spp. were enriched with vaginal deliveries. However, this is also an organism that has been previously implicated in intra-uterine infection and adverse outcomes.

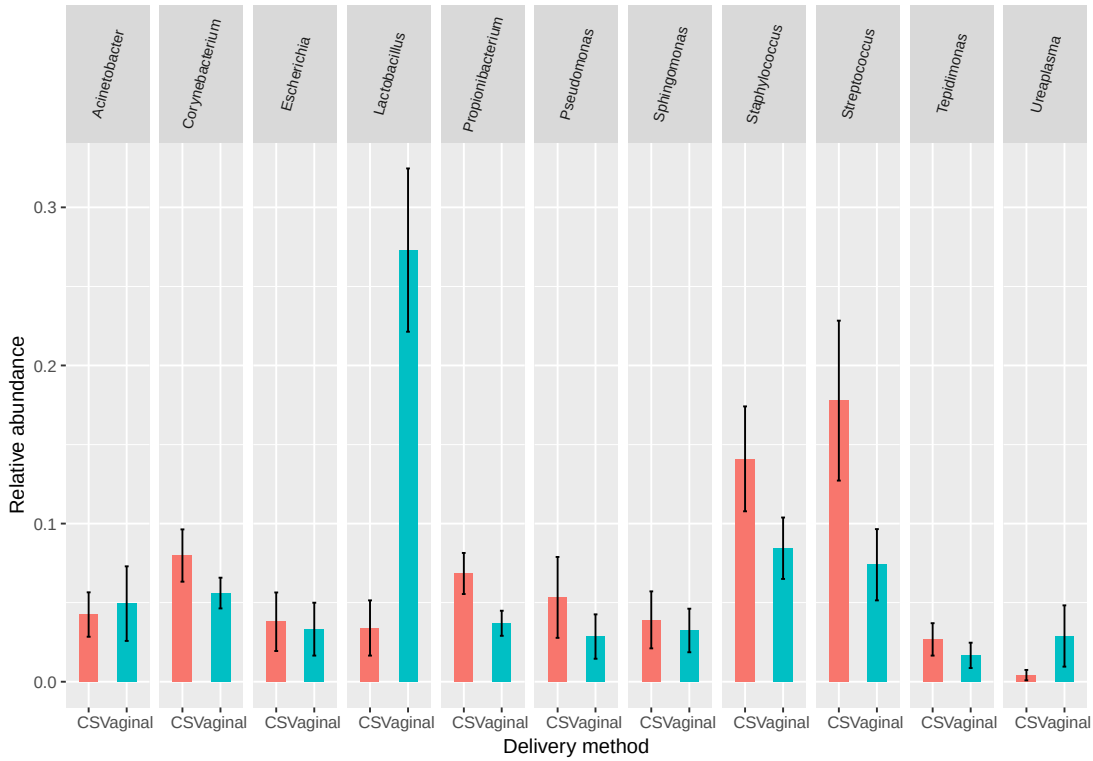


Figure 6.6 – The top 10 genera from CS (red) and vaginal (blue) deliveries with highest mean relative abundance. Genera were combined to form 11 unique groups and the mean relative abundances between the two groups were compared. Error bars show 95% CI around the mean.

Table 6.7 – Difference in mean relative abundance of the 11 CS and vaginal genera with highest mean relative abundances. Student's T-test and two-sided test for significance used for comparison of means (*<0.05).

Genus	Difference in % relative abundance (CS-vaginal)	T	P
Streptococcus	10.38	3.67	3.24E-04*
Staphylococcus	5.65	2.88	4.42E-03*
Propionibacterium	3.15	4.05	7.45E-05*
Pseudomonas	2.48	1.66	9.80E-02
Corynebacterium	2.37	2.42	1.63E-02*
Tepidimonas	1.01	1.52	1.30E-01
Sphingomonas	0.67	0.58	5.64E-01
Escherichia	0.47	0.37	7.14E-01
Acinetobacter	-0.69	-0.49	6.24E-01
Ureaplasma	-2.48	-2.47	1.45E-02*
Lactobacillus	-23.9	-8.59	3.57E-15*

6.3.4 Differential abundance analyses with outcomes of interest

Following further filtering and transformation, as outlined in Section 6.2.5.1, OTUs from 249 identified genera were retained, and their relative abundances compared between the three pregnancy outcome groups. Multi-level models were first run unadjusted. Models were then run again with adjustment for fixed level potential confounders.

6.3.4.1 Differentially abundant OTUs and delivery method

When OTU abundances were compared between vaginally and CS delivered placenta, of the top 46 most differentially abundant OTUs (P<0.01), 30 mapped to *Lactobacilli* spp. (see Appendix Table A 16, page 261). These data provide further support for the observations discussed in section 6.3.3.

6.3.4.2 Identification of taxa enriched in sPTB placental samples

The primary hypothesis of this study was that certain organisms would be differentially abundant in placental tissue, according to pregnancy outcome. Specifically, it was hypothesised that an overgrowth or colonisation of particular organisms would lead to a pathogenic inflammatory response, followed by early birth. Therefore, it was important to identify organisms present at higher abundances in the placental tissue of sPTB pregnancies, compared to term or nsPTB pregnancies. It is possible that certain placental organisms are protective against sPTB and would therefore be less abundant in sPTB placenta, perhaps equivalent to the role of *Lactobacilli* in vaginal health. However, this hypothesis will not be explored here. In order to explore both higher and lower levels of taxonomic assignment, differential abundance testing was carried out at the OTU and genus level. Comparisons were made between sPTB placenta and nsPTB placenta, as well as sPTB and term placenta.

6.3.4.2.1 Unadjusted models

Table 6.8 lists those genera that were found to be at a greater abundance in sPTB placenta (P<0.01). As will be discussed below, a number of these genera, such as *Ureaplasma* and *Mycoplasma,* have been consistently associated with adverse pregnancy outcomes in previous studies. Others, such as *Tepidimonas*, have not.

Table 6.8 – Genera enriched in sPTB placenta using an unadjusted model, P<0.01.

Genera enriched versus nsPTB	Genera enriched versus term
Ureaplasma	*Tepidimonas*
Mycoplasma	*Mycoplasma*
Salinicoccus	*Salinicoccus*
Prevotella	*Capnocytophaga*
Bilophila	*Erwinia*
	Bilophila
	Finegoldia
	Neisseria
	Sneathia
	Coprobacillus
	Phenylobacterium
	Actinobaculum

6.3.4.2.2 Enriched OTUs and genera in sPTB versus nsPTB placenta

All OTUs with higher abundance in sPTB versus nsPTB placental tissue, and with a P<0.01 after adjustment for confounders, are listed in Table 6.9. When this comparison was repeated at the level of genus, *Ureaplasma*, *Mycoplasma*, and *Capnocytophaga* were found to be at a significantly higher abundance at a threshold of P<0.01 (Table 6.10). It is important to note that when P-values were adjusted using FDR, none of the comparisons fell below the widely used threshold of Q<0.1. This may to some extent reflect the large number of tests carried out in these comparisons. When only parenchyma tissue was considered, only *Capnocytophaga* was higher in sPTB versus nsPTB placenta and had a P<0.01 (data not shown).

Table 6.9 – OTUs enriched in sPTB vs nsPTB placenta with P<0.01 using an unadjusted Limma model.

OTU	Genus	Species from NCBI BLAST	Log2 fold change (95% CI)	P	Q
884685	*Finegoldia*	-	0.24 (0.06-0.43)	0.0093	0.41
518865	*Fusobacterium*	*nucleatum*	0.2 (0.06-0.34)	0.00469	0.34
531206	*Mogibacterium*	-	0.46 (0.12-0.8)	0.00791	0.41
317108	*Mycoplasma*	*hominis*	0.83 (0.31-1.35)	0.00197	0.27
548556	*Mycoplasma*	*hominis*	1.03 (0.35-1.7)	0.00326	0.33
573950	*Mycoplasma*	*hominis*	0.32 (0.08-0.55)	0.00779	0.41
575236	*Mycoplasma*	*hominis*	0.65 (0.23-1.06)	0.00253	0.31
594001	*Mycoplasma*	*hominis*	0.5 (0.2-0.79)	0.00102	0.27
OTU4796	*Mycoplasma*	*hominis*	0.2 (0.06-0.34)	0.00427	0.33
OTU47	*Mycoplasma*	*hominis*	0.56 (0.17-0.95)	0.00546	0.38
NROTU83	*Mycoplasma*	*hominis*	0.33 (0.08-0.58)	0.00886	0.41
183158	*Peptoniphilus*	-	0.25 (0.09-0.4)	0.00194	0.27
4332287	*Streptococcus*	*sanguinis*	0.35 (0.15-0.56)	0.0008	0.27
15806	*Ureaplasma*	*parvum/ urealyticum*	1.99 (0.86-3.12)	0.00064	0.27

Table 6.10 – Genera enriched in sPTB vs nsPTB placenta with P<0.01 using an unadjusted Limma model.

Genus	Log2 fold change (95% CI)	P	Q
Ureaplasma	2.4 (0.96-3.85)	0.0012	0.25
Mycoplasma	1.58 (0.58-2.59)	0.0022	0.25
Capnocytophaga	0.6 (0.2-0.99)	0.0031	0.25

6.3.4.2.3 Enriched OTUs and genera in sPTB versus term placenta

When OTU abundance was compared between sPTB and term delivered placenta, 47 OTUs were more abundant in the sPTB cohort with a P<0.01 (Table 6.11). As seen above, many of these OTUs had duplicate species assignments. For example, 5 OTUs mapping to *Mycoplasma hominis* assignments were significantly enriched in sPTB placenta. This repetition may reflect the presence of a number of different sub-species within the placenta, or redundancy in the OTU clustering method used.

Table 6.11 – OTUs enriched in sPTB versus term placenta following adjustment, P<0.01.

OTU	Genus	Species from NCBI BLAST	Log2 fold change (95% CI)	P	Q
574249	*Actinobaculum*	*schaalii*	0.32 (0.09-0.56)	0.00603	0.19
1077793	*Alloiococcus*	-	0.45 (0.19-0.7)	0.00065	0.12
419366	*Anaerococcus*	-	0.37 (0.09-0.65)	0.0097	0.24
503387	*Anaerococcus*	-	0.31 (0.1-0.52)	0.0036	0.15
1002634	*Anaerococcus*	-	0.2 (0.05-0.35)	0.00798	0.21
276484	*Bacteroides*	-	0.13 (0.04-0.23)	0.00427	0.16
2066505	*Bacteroides*	-	0.09 (0.03-0.15)	0.00285	0.13
2283111	*Bacteroides*	*uniformis*	0.1 (0.04-0.15)	0.0009	0.12
360660	*Blautia*	-	0.26 (0.12-0.4)	0.0003	0.12
NCROTU4734	*Blautia*	-	0.13 (0.04-0.22)	0.00648	0.19
71957	*Capnocytophaga*	-	0.21 (0.07-0.35)	0.00438	0.16
1010329	*Capnocytophaga*	*gingivalis*	0.45 (0.18-0.72)	0.00111	0.12
4354247	*Capnocytophaga*	-	0.15 (0.05-0.25)	0.00528	0.18
NCROTU55	*Capnocytophaga*	-	0.19 (0.06-0.32)	0.00432	0.16
189459	*Clostridium*	-	0.23 (0.11-0.36)	0.00036	0.12
369763	*Coprobacillus*	-	0.14 (0.05-0.23)	0.00222	0.12
308983	*Coprococcus*	-	0.15 (0.06-0.25)	0.00217	0.12
977739	*Corynebacterium*	-	0.15 (0.04-0.25)	0.00628	0.19
1010113	*Erwinia*	-	0.29 (0.1-0.48)	0.0025	0.12
317108	*Mycoplasma*	*hominis*	0.67 (0.25-1.08)	0.00189	0.12
548556	*Mycoplasma*	*hominis*	0.89 (0.35-1.44)	0.00133	0.12
575236	*Mycoplasma*	*hominis*	0.5 (0.17-0.83)	0.00336	0.15
NCROTU4796	*Mycoplasma*	*hominis*	0.15 (0.04-0.26)	0.00855	0.22
NROTU47	*Mycoplasma*	*hominis*	0.46 (0.15-0.77)	0.00427	0.16
364341	*Oscillospira*	-	0.16 (0.06-0.26)	0.00216	0.12
548692	*Oscillospira*	-	0.14 (0.05-0.22)	0.0022	0.12
291090	*Parabacteroides*	-	0.09 (0.03-0.15)	0.00244	0.12
254401	*Phenylobacterium*	-	0.08 (0.02-0.14)	0.00623	0.19
262857	*Salinicoccus*	-	0.32 (0.13-0.5)	0.00099	0.12
931528	*Sporobacterium*	-	0.11 (0.03-0.18)	0.00545	0.18
526131	*Streptococcus*	*pneumonia*	0.45 (0.18-0.72)	0.00101	0.12
859700	*Streptococcus*	-	0.5 (0.19-0.81)	0.0016	0.12
942927	*Streptococcus*	-	0.38 (0.19-0.58)	0.00016	0.12
983335	*Streptococcus*	-	0.52 (0.21-0.84)	0.00121	0.12
986708	*Streptococcus*	*sanguinis*	0.23 (0.09-0.38)	0.00207	0.12
996434	*Streptococcus*	-	0.51 (0.2-0.82)	0.00141	0.12
1023716	*Streptococcus*	*oralis/mitis*	0.3 (0.1-0.5)	0.00405	0.16
1061897	*Streptococcus*	-	0.57 (0.21-0.93)	0.00196	0.12
1078207	*Streptococcus*	-	0.37 (0.1-0.65)	0.00766	0.2
4295788	*Streptococcus*	*sanguinis*	0.19 (0.05-0.33)	0.00726	0.2
4321400	*Streptococcus*	*pneumonia*	0.52 (0.22-0.83)	0.00092	0.12
4332287	*Streptococcus*	*sanguinis*	0.28 (0.11-0.44)	0.001	0.12
4433192	*Streptococcus*	-	0.31 (0.09-0.53)	0.00545	0.18
189076	*Sutterella*	-	0.13 (0.05-0.22)	0.00278	0.13
697479	*Tepidimonas*	*aquatica*	1.73 (0.74-2.71)	0.00066	0.12
15806	*Ureaplasma*	*parvum/ urealyticum*	1.24 (0.34-2.14)	0.0073	0.2
676611	*Ureaplasma*	*parvum/ urealyticum*	0.23 (0.07-0.39)	0.00466	0.16

Table 6.12 – Genera enriched in sPTB vs term placenta following adjustment, P<0.01.

Genus	Log2 fold change (95% CI)	P	Q
*Capnocytophaga**	0.67 (0.36-0.99)	3.73E-05	0.01
*Tepidimonas**	2.42 (1.26-3.58)	5.44E-05	0.01
*Salinicoccus**	0.36 (0.13-0.6)	2.40E-03	0.17
*Coprobacillus**	0.12 (0.04-0.2)	2.68E-03	0.17
Mycoplasma	1.17 (0.37-1.98)	4.42E-03	0.22
*Bilophila**	0.15 (0.04-0.25)	5.92E-03	0.25
*Erwinia**	1.2 (0.32-2.08)	7.71E-03	0.25
*Actinobaculum**	0.34 (0.08-0.6)	9.99E-03	0.25

*also <0.01 in parenchyma only analysis

The majority of OTUs and genera groupings did not reach statistical significance, once adjusted for multiple testing. However, a number of the associations were with taxa that have previously been associated with sPTB, such as *Mycoplasma* spp. and *Ureaplasma* spp. The observation that certain associations replicated previous findings in the field served to support the reliability of our data, even if significance was lost following adjustment for multiple testing. Therefore, it was considered reasonable to further explore the nature of the organisms associated with sPTB across all the tests carried out. To do this, the characteristics of genera associated with sPTB via the OTU or genus level analyses at a significance of P<0.01 were investigated further. The core habitats of these 26 genera of interest, and relevant references for their involvement in adverse pregnancy outcomes, are summarised in Table 6.13.

The associations identified via these analyses could reflect both widespread differences in abundance, which are largely consistent between groups, or less common 'outlying' cases of extreme and highly abundant infections. To explore the nature of these associations further, genus relative abundances were calculated for individual pregnancies, by pooling biological replicates into one sample. Plots were then constructed with relative abundance on the Y-axis and pregnancies on the X-axis. Separate plots for sPTB, term, and nsPTB cases were compared in order to visually explore if abundance patterns were largely consistent across groups, or associations were mainly driven by individual, influential cases (see Appendix Figure A 1 – A 3 pages 262-264 for all plots).

Table 6.13 – Normal habitat and relevant references relating to PTB for all genera enriched in sPTB placenta in our cohort with P <0.01 in either OTU or genus level analyses.

Phylum	Genus	Species isolated from human…	Previous associations with preterm birth
Actinobacteria	*Actinobaculum*	Blood, urine, oral cavity	
	Corynebacterium	Oral cavity, mucosa and skin	Oh et al. (2010)
Bacteroidetes	*Bacteroides*	Vagina, mouth, gastro-intestinal tract	Jones et al. (2009), (Han et al., 2009)
	Capnocytophaga	Oral cavity	Ernest and Wasilauskas (1985), (Lopez et al., 2010a)
	Parabacteroides	Faeces	
Firmicutes	*Alloiococcus*	Vagina, oral cavity, auditory canal	
	Anaerococcus	Vagina, oral cavity, preterm gut, skin	
	Blautia	Gastro-intestinal, preterm gut	
	Clostridium	Vagina, gastro-intestinal	
	Coprobacillus	Gastro-intestinal	DiGiulio et al. (2010)
	Coprococcus	Gastro-intestinal	
	Finegoldia	Vagina, oral cavity, and skin	
	Mogibacterium	Oral cavity	
	Oscillospira	Gastro-intestinal	
	Peptoniphilus	Vagina, mouth, gastro-intestinal	Jones et al. (2009), (Wang et al., 2013)
	Salinicoccus	-	
	Sporobacterium	-	
	Streptococcus	Vagina, mouth, skin, gastro-intestinal	(Jones et al., 2009, Oh et al., 2010)
Fusobacteria	*Fusobacterium*	Oral cavity	Combs et al. (2014); DiGiulio et al. (2010); Gardella et al. (2004); Han et al. (2009); Jones et al. (2009); Wang et al. (2013)
	Bilophila	Faeces, oral cavity, vagina	
	Erwinia	Meconium	
	Phenylobacterium	-	
	Sutterella	Vagina, gastro-intestinal	Ollberding et al. (2016)
	Tepidimonas	-	
Tenericutes	*Mycoplasma*	Vagina, oral cavity	DiGiulio et al. (2010); Hitti et al. (1997); Han et al. (2009); Jones et al. (2009); Marconi et al. (2011); Onderdonk et al. (2008a); Wang et al. (2013)
	Ureaplasma	Vagina	Combs et al. (2014); DiGiulio et al. (2010); Han et al. (2009); Jalava et al. (1996); Jones et al. (2009); Wang et al. (2013)

Figure 6.7 shows a selection of these individual pregnancy plots, demonstrating patterns of interest. *Erwinia, Fusobacteria*, and *Mycoplasma* were all consistently identified as being at higher abundance in sPTB cases compared to the other groups, using Limma regression models. It is clear from the plots that these associations were driven almost entirely by 5 pregnancies in which the majority of reads mapped to these three genera. This observation supports a scenario in which a single organism came to dominate the placenta of these specific pregnancies, potentially triggering the sPTB event. For these three organisms, little evidence of their presence in term or nsPTB placenta was found. This picture is not so clear for other genera, which tended to be more consistently present across samples, generally at lower abundances.

Ureaplasma enrichment within sPTB placental tissue was more apparent in statistical comparisons with nsPTB, as opposed to term placenta. A substantial number of term placenta had a high abundance of *Ureaplasma*, as did a number of sPTB births. However, only one read from *Ureaplasma* mapped to nsPTB placenta. Since *Ureaplasma* is a common vaginal organism, and a lower proportion of nsPTB deliveries were vaginal, this may account for at least some of this observed pattern. The distributions of reads mapping to other genera that were at higher abundances in sPTB placenta, such as *Peptoniphilus, Capnocytophaga*, and *Finegoldia,* were also not as exclusively concentrated in those of the sPTB group, as the first three genera discussed. These genera were generally found at lower abundances (<10%), but in a larger number of individual placentas.

6.3.4.3 Cervical suture

Only 11 samples were available with cervical suture data, following filtering. Therefore, an investigation of the types of organisms present in placenta following cervical suture, was not carried out within the sequencing cohort.

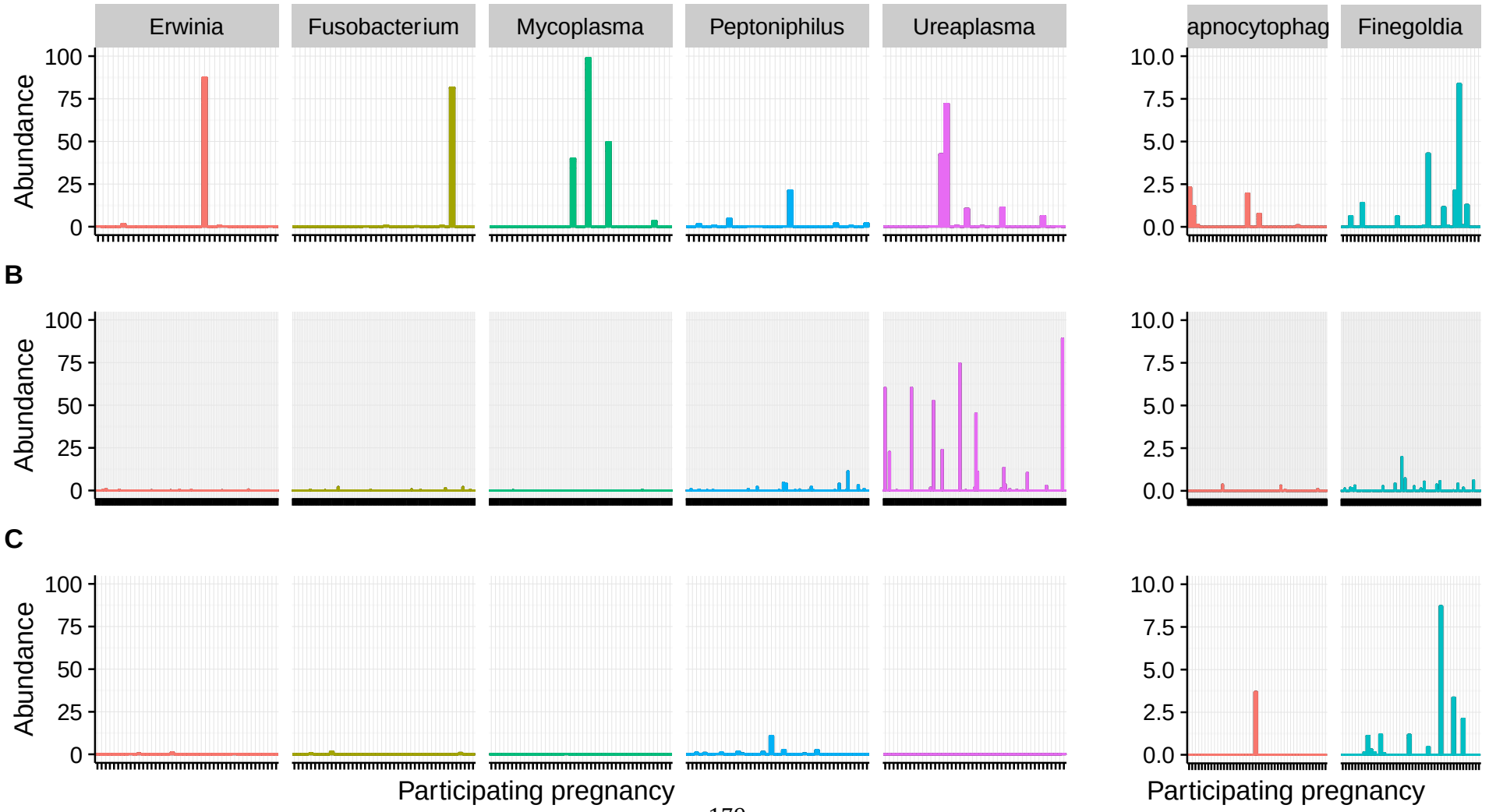


Figure 6.7 – Abundance (%) of total reads in placentas from individual pregnancies for 6 genera of interest. A) sPTB placenta B) Term delivered placenta C) nsPTB placenta.

6.3.5 Between sample (beta) diversity

Weighted UniFrac, unweighted UniFrac, and Bray-Curtis distance matrices were produced from the data to estimate beta diversity. Distances were then plotted using PCoA to visualise the first two axes. They were coloured by pregnancy outcome (Figure 6.8). Samples did not clearly cluster by pregnancy outcome with any of the methods used. Therefore, in order to explore the distance matrices in a quantitative manner, which may be more sensitive to underlying patterns, beta diversity was compared between groups in both adjusted and unadjusted analyses using adonis.

Adonis does not allow for the addition of a random intercept to make a multi-level model. As a result, using all samples at once would not account for likely correlations between biological replicates. Therefore, two analysis designs were used. The first used all samples with the knowledge that there might be a higher false positive rate. The second merged all biological replicates from individual pregnancies together, producing a more conservative estimate of associations.

Results from all 12 adjusted and unadjusted tests produced very similar R^2 values, all of which were significant at P=0.001 (Table 6.14). Despite this significance, it is clear that only a very small proportion of the variance between samples (~2%) is accounted for by the three outcome groups of interest: nsPTB, sPTB, and term delivery. It is reasonable to conclude that these very subtle differences in overall taxonomic composition are unlikely to be clinically relevant. Evidence from these analyses does not support the existence of a meaningful 'preterm placental microbiome'.

Table 6.14 – R^2 values for comparisons of three diversity metrics between three groups of interest (sPTB, nsPTB, term) for adjusted and unadjusted models using all samples or data merged for individual pregnancies.

Beta diversity metric	All samples (R^2)		Participant level (R^2)	
	Unadjusted 3 group model	Adjusted 3 group model	Unadjusted 3 group model	Adjusted 3 group model
Weighted UniFrac	0.019*	0.017*	0.0021*	0.022*
Unweighted UniFrac	0.015*	0.017*	0.0016*	0.018*
Bray-Curtis	0.018*	0.021*	0.0018*	0.020*

*P=0.001

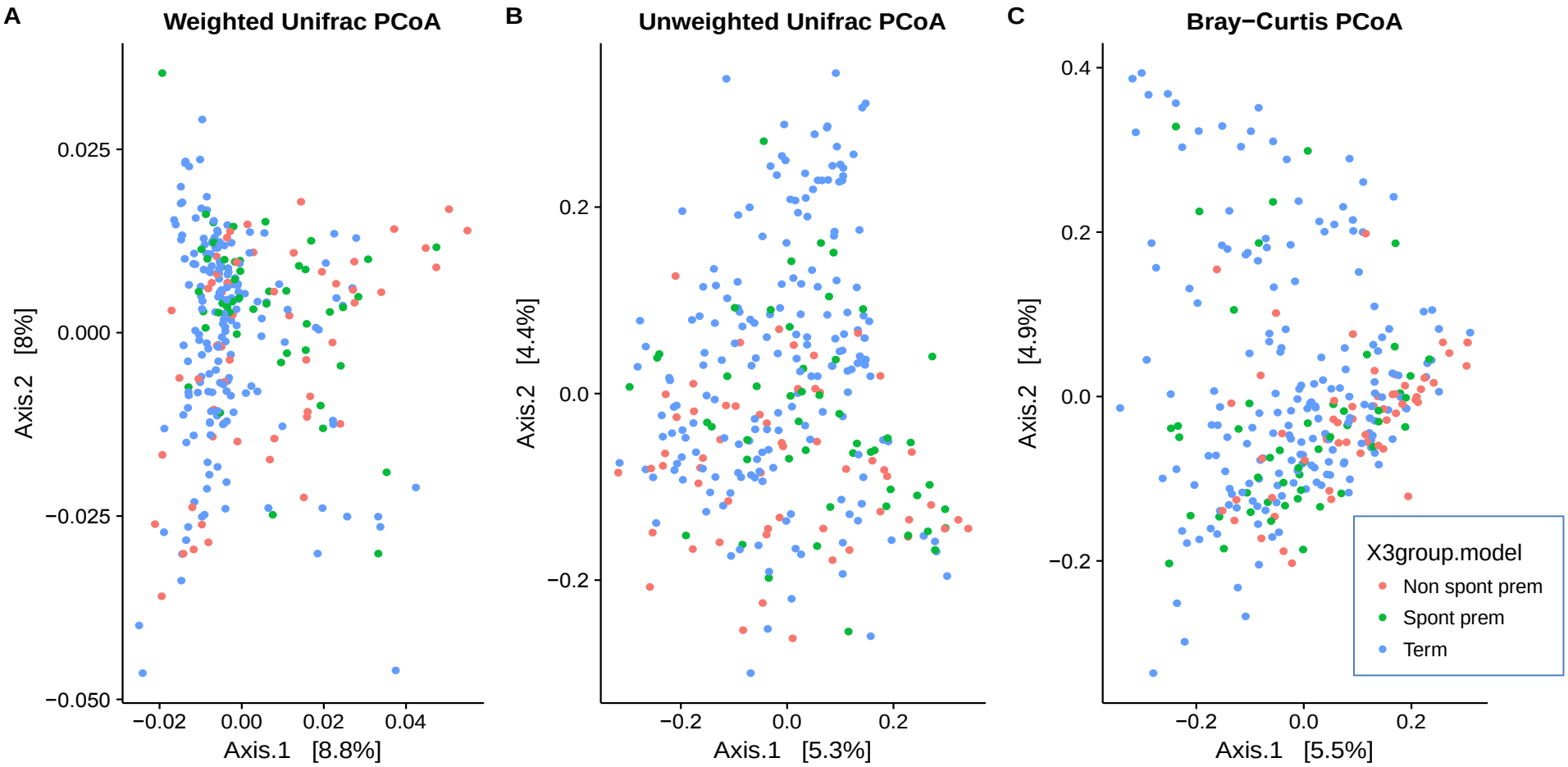


Figure 6.8 – PCoA plot of first two axes from A) weighted UniFrac; B) unweighted UniFrac; C) Bray-Curtis distances using VST normalised counts of all samples in the cohort.

6.3.6 Within sample (alpha) diversity

Within sample, alpha diversity can be quantified using a number of indices. These tend to calculate a score based on the total number of species/OTUs present within a sample (richness), and, sometimes, evenness of read distributions between these groups. Table 6.15 summarises estimates of Richness and Shannon indices across a subset of the whole cohort, rarefied to 1000 reads and averaged over 100 iterations. A total of 269/321 samples were retained after rarefaction. The mean number of OTUs in any one sample was 48.5. As discussed above, this is likely to be an inflation of the true species richness within samples. This potential inflation is a result of errors in the experimental and computational processes that create these sequencing cohorts, which serve to introduce false variation into the nucleotide sequences.

Table 6.15 – Summary of Shannon diversity and total richness estimates for whole cohort.

Index	Mean (SD)	Median (IQR)	Range
Shannon index	2.3 (0.95)	2.5 (1.7-3.0)	0.05-4.5
Total richness	48.5 (25.4)	43.6 (31.7-48.5)	5.1-169.2

As demonstrated in Figure 6.7, placental infection within certain sPTB samples appeared to be driven by growth of a single organism. It follows that placenta with more significant infections, as reflected by higher bacterial loads, may then have lower alpha diversity measures. To explore this, total bacterial load (see Chapter 5), was plotted against the Shannon index for each individual sample. The Shannon index increases with the increasing evenness and richness of a community.

Figure 6.9 shows that as 16S load increases, mean Shannon index decreases. This implies that a larger bacterial presence in placental tissue is associated with a reduction in alpha diversity. This reduction in diversity is both in terms of evenness and richness. This scenario could be interpreted as an expansion of one organism at the expense of many others. When a linear regression model is fitted to this relationship, for every 1 increase in 16S copy number, there is a -0.24 reduction in Shannon index (P=0.00017). A similar, significant observation was noted when Simpsons index was used instead of Shannon in this model (data not shown). No relationship was observed when mean richness was used instead of the Shannon value.

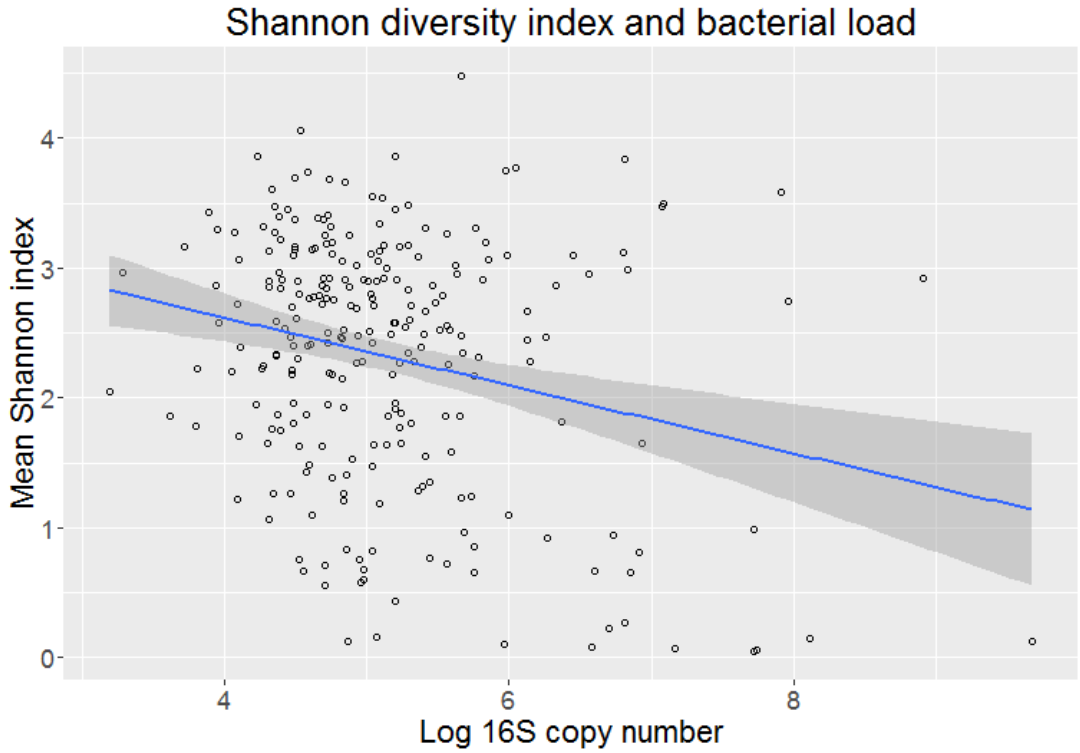


Figure 6.9 – The mean Shannon index reduces with increasing bacterial load in placental tissue from a rarefied cohort.

6.3.6.1 Alpha diversity and pregnancy outcome

To investigate associations between alpha diversity and pregnancy outcome, multi-level regression models were run, with participant as a random intercept. If a significant proportion of infectious mediated preterm births within the cohort were driven by overgrowth of a single, or a small number of species, it would be reasonable to hypothesise that sPTB may be associated with a reduction in alpha diversity. However, no significant difference was observed in the mean Shannon diversity index between sPTB and nsPTB (P=0.53, adjusted model), or between sPTB and term (P=0.14, adjusted model) placenta, in univariate analyses, as well as those adjusted for the potential effects of delivery method, ethnicity, BMI, smoking, and tissue type. There was also no association observed when mean richness was the outcome (P=0.29, P=0.39, for adjusted comparisons between sPTB and nsPTB or term placenta, respectively). Delivery method was not associated with any difference in Shannon diversity (P=0.96, adjusted model) or total richness (P=0.1, adjusted model).

6.4 Discussion

This study has shown that bacterial DNA from a wide variety of organisms was present in the placenta of both normal and complicated pregnancies. In order to understand the clinical significance of such observations, identifying where the sequenced DNA originates from is critical. This will facilitate the development of clinical interventions to mitigate the adverse impacts of infection during pregnancy. From the analyses conducted in this chapter, it is clear that delivery method contributed significantly to the taxonomic composition of the sequenced placental samples. Nevertheless, a number of organisms were clearly enriched within sPTB placental samples, independently of mode of delivery.

Analyses focusing on diversity metrics also identified some associations of interest. Shannon diversity was negatively associated with bacterial load, implying that those placental samples that were most 'infected', were also less diverse and less evenly colonised. This perhaps represents overgrowth of a single or small number of species in a small number of extreme cases. Despite this, alpha diversity did not associate with pregnancy outcome specifically, and beta diversity measures differed minimally between groups. These observations together, imply that it is largely in the presence of single or small numbers of organisms of interest that risk of sPTB is increased. Evidence for the existence of a shared 'preterm placental microbiome' was less conclusive. A summary of the key findings from this chapter are as follows:

1. The majority of organisms recovered from the placenta were assigned to five phyla: *Firmicutes, Proteobacteria*, *Tenericutes, Bacteroidetes* and *Actinobacteria.*

2. Relative abundances of some of the most prevalent genera varied by delivery method. Vaginally delivered placenta were dominated by *Lactobacilli* spp. and CS deliveries by common skin commensals.

3. A number of OTUs and genera were consistently enriched in sPTB placenta, some of which are well known organisms hypothesised to be involved in sPTB, although few reached statistical significance when adjusted for multiple testing.

4. For some enriched organisms, the associations appear to be largely driven by highly infected, 'outlying' placental infections of single pregnancies.

5. Using three measures of beta diversity, the placental microbiome did not cluster by outcome group. Although statistical tests of difference were significant, the proportion of variance accounted for by the model was very low (<2%) and thus assumed to be of limited clinical significance.

6. With increasing bacterial load, evenness and richness was reduced within samples, as quantified by the Shannon index.

7. Alpha diversity did not associate with the pregnancy outcomes of interest.

Aagaard et al (2014) described a somewhat similar high-level taxonomic community in their microbiome analysis of placental parenchyma samples as those found in this study. Three of the top four most abundant phyla in their study were in the top five most abundant in this cohort (*Bacteroidetes, Firmicutes,* and *Tenericutes)*. The five most abundant phyla found in our cohort contain many well-known human commensals, from the gut, to the skin, oral, and vaginal microbiomes (Huttenhower et al., 2012).

Many of the organisms recovered from our placental samples are also known vaginal and skin commensals, such as *Lactobacilli* and *Corynebacterium* spp. respectively. It was shown that some of these most abundant organisms did associate strongly with delivery method. These patterns were also observed in parenchyma tissue, which was collected from beneath the protective fetal membrane. Similar observations were drawn from villous samples that were analysed separately. Given the hypothesis that at least some intra-uterine colonisation will be the result of ascending vaginal infection, it would be overly conservative to label all vaginal species, or indeed all skin-associated organisms, as delivery contaminants. Indeed, none of the delivery-associated organisms were exclusive to one or other delivery type. However, these associations do make it difficult to describe conclusively a placental microbiome, which is distinct from the 'delivery microbiome'. Further, the observation of a low-level, relatively diverse placental microbial signature is supported by other recent molecular studies working in the same tissues (Doyle et al., 2014, Jones et al., 2009, Aagaard et al., 2014).

The most powerful observations from this cohort surround comparisons of normalised abundances of OTU and genus level counts, between sPTB and other pregnancy outcomes. The most significantly enriched genera and OTUs were, in large part, organisms previously associated with adverse pregnancy outcomes. Although few of

these associations survived conservative cut-offs following multiple correction. Some, such as *Mycoplasma*, *Ureaplasma*, and *Fusobacteria* spp., have been known as opportunistic intra-uterine pathogens for decades (Combs et al., 2014, DiGiulio, 2012, Jones et al., 2009, Hitti et al., 1997). Others, such as *Capnocytophaga*, are less well studied with respect to sPTB pathogenesis. Interestingly, the organisms that associated with sPTB in the Aagaard et al. (2014) study, such as *Burkholderia* spp., did not overlap with those found here. A smaller study describing the placental membrane microbiome from preterm and term pregnancies reported results that were much more similar to our own (Doyle et al., 2014). For example *Fusobacteria*, *Mycoplasma*, and *Ureaplasma* spp. were also enriched in their vaginally delivered sPTB samples, when compared to term placenta.

There was a notable prevalence of oral bacteria in the organisms enriched in sPTB placenta within this cohort. These associations contribute to a growing body of research supporting a link between organisms isolated from the placental and the oral cavities. These associations are supported by both human and animal data. Han et al. (2004) injected *F. nucleatum* into the tail of pregnant mice, ultimately inducing preterm delivery and fetal loss. These live *F. nucleatum* cells were found to have proliferated within uterine tissues examined from these animals, but they did not persist in other organs such as the spleen or liver. As previously discussed, Aagaard et al. (2014) also reported a significant overlap of their data with the oral rather than vaginal HMP dataset. Indeed, from our data it was clear that at least one sPTB placenta was significantly infected with *Fusobacterium* spp. This was a relatively early, 29th week delivery by CS, in which the infant was admitted to intensive care. This type of single organism overgrowth, in which a very large proportion of reads mapped to a single genus, was also observed in at least two other instances for *Mycoplasma hominis* and, curiously, OTUs mapping to the *Erwinia* genus. *Erwinia* spp. were present in four placental samples from the same pregnancy (two technical replicates of parenchyma, one more parenchyma biological replicate, and one villous sample) at between 86 and 89% abundance. The abundances were very similar between all four samples from the same pregnancy, demonstrating the robustness of this observation. Furthermore, the placenta was taken from a very early vaginal birth at 24 weeks: this infant also went to the NICU. *Erwinia* species have not been previously associated with sPTB. Indeed it is a genus of bacteria containing mostly plant pathogens, which makes this observation somewhat unexpected. However, it has previously been isolated and cultured from

meconium samples from a number of very preterm neonates (Moles et al., 2013). This unexpected but interesting association therefore requires further exploration.

Capnocytophaga, was present across a number of samples at a relatively low total abundance, when compared to genera such as *Mycoplasma.* However, it was one of the two genera that remained significantly associated with sPTB placenta following adjustment for multiple testing. This anaerobic organism is usually isolated from the oral cavity, and rarely the genital tract. *Capnocytophaga* infections tend to be most commonly reported in immunocompromised children and are known to be involved in periodontitis (Campbell and Edwards, 1991). The genus has been previously associated with intra-uterine infection in a handful of publications (Douvier et al., 1999, Lopez et al., 2010b, Hill, 1998). In a recent review of the literature, Lopez et al. (2010b) identified a total of 19 reported cases of *Capnocytophaga* infection of the amniotic cavity and/or a preterm infant and concluded that:

> *'Infection with* Capnocytophaga *sp. is therefore probably an underestimated aetiology of occult chorioamnionitis and preterm delivery leading to neonatal infection.'*

Capnocytophaga's occasional, but independently reported, associations with adverse pregnancy outcomes, as well as its status as an oral opportunistic pathogen, provide support for its association with sPTB in our cohort. Furthermore, it is often in immunocompromised patients that infections of this organism take hold. This is interesting, considering the immuno-suppressed state of pregnancy. Taken together, this may well be a promising and underappreciated candidate for placental mediated sPTB pathogenesis. *Capnocytophaga* was also isolated from one nsPTB placenta in our cohort. The infant associated with this placenta was delivered by CS at 29 weeks. It was significantly growth restricted, with a birthweight of only 634g, and was admitted to intensive care. This mother had previously delivered another preterm infant. Whether this *Capnocytophaga* infection contributed in some way to the fetal growth of this infant is worthy of further exploration.

The significant enrichment of *Tepidimonas* within sPTB versus term placenta, and its particularly high log2 fold change value, is perhaps the most enigmatic result from these analyses. Over 12,000 reads mapped to 5 *Tepidimonas* labelled OTUs. This was found across all outcome groups and many individual pregnancies. A literature search for publications relating to human infections and *Tepidimonas* spp. revealed only one

case, where there was a bone marrow infection in a leukemia patient (Ko et al., 2005). This lack of previous human associated evidence questions the validity of the widespread presence of this genus in our cohort. This genus was first described as an isolate from a hot spring in central Portugal (Moreira et al., 2000). Thus this observation may be a false positive association between outcome and a contaminating genus that is specific to our cohort.

Bacteroides, Peptoniphilus, Streptococcus, and *Suterella* species were some of the most significantly enriched genera and OTUs within our sPTB cohort. These genera have also previously been associated with preterm birth (Combs et al., 2014, Wang et al., 2013, Hitti et al., 1997). In one sample with a particularly high *Peptoniphilus* abundance, reads mapping to *Sneathia*, another well-known sPTB associated organism, made up over 25% of total reads. Some genera, such as *Mogibacterium* and *Finegoldia* contain oral and vaginal species respectively, with the former notably associated with periodontitis and the latter with BV (Nakazawa et al., 2000, MacIntyre et al., 2015). These initial associations require further exploration for validation.

These data support the notion that certain bacteria, when present in the placenta, may contribute to an inflammatory response and a subsequent preterm delivery. However, the data do not support the existence of a 'unique preterm microbiome' in terms of a structured community, which is shared amongst this particular obstetric group, and distinguishable from other outcome groups. This interpretation is supported both by the graphical ordination of the beta diversity data, and the quantitative statistical tests that were conducted. Although the P-value of all adonis tests could lead us to reject the null hypothesis of no difference in community structure between groups, only around 2% of variation in the beta diversity matrices was ever accounted for by these groupings. This is a very small proportion and unlikely to be of clinical relevance. Given the data at hand, it is reasonable to assert that the high-level community structure of the placental microbiome in this cohort is not significantly associated with pregnancy outcome.

There were a number of limitations to this study. The first was that we were unable to sequence the whole cohort for which qPCR data were available, necessarily limiting power to identify all associations of interest. This will hopefully be addressed in the future when more resources become available. Despite this, our study provides one of the largest cohorts of 16S sequenced placental tissue from sPTBs in the literature.

Furthermore, some of these cases were delivered at extremely low GAs, making this a particularly valuable and rare sample set.

A further limitation of this study was that some clinical categories of interest, such as delivery method or maternal smoking status, were incomplete. This limited the size of some of the final adjusted models. It would also have been of interest to examine the association of microbiome profiles with the well-known clinical phenotype of chorioamnionitis, but this data was unavailable in the BBB clinical records. In the future it would also be of great interest to collect parallel samples from other maternal body sites. This could help to identify the origins of organisms of interest.

Sequencing a tissue of low biomass is a challenge for reasons discussed throughout this PhD. Those microbiome studies focusing on tissues of high biomass, such as the vagina, oral cavity, or intestine, produce 16S sequencing outputs of even and high per-sample read depth. This was not the case for our placental cohort. Although for some samples very large numbers of reads were procured, the likely underlying variation in starting material within this tissue meant that normalisation of template amount was a particular challenge for this cohort. As a result, very few reads were produced for certain samples and there was a very wide range of sequencing depths across the cohort. This presented challenges for statistical interpretation of the data, which were met via normalisation methods originally designed for RNA-seq and micro-array data. Additionally, 73 samples (19%) were discarded during the filtering process, as their read depth, following filtering, was not considered sufficient to be representative of the entire diversity of the population sampled. The development of techniques, some of which are already starting to be used in which microbial DNA enrichment precedes sequencing, will be a powerful improvement in future studies.

The dissection and extraction procedures for this study were carried out using rigorously controlled, sterile procedures. However, the samples were not originally collected with the intention to be used in microbial analyses. Placenta were not collected under strictly sterile conditions and it is acknowledged that this may limit the reliability of observations noted here. Similarly, the storage and cleaning reagents, *RNAlater* and PBS, which were used by the collection team were not available to use as comparative sequencing controls, in the same way that extraction reagents were. Perhaps this explains the unexpected associations with genera such as *Tepidimonas*.

However, adjustment in final models for hospital collection site, will have accounted for at least some site-specific contamination patterns.

The specific study of the placental microbiome remains in its infancy. This research and other similar data suggest that there may be a low-level non-pathogenic placental microbiome present in many, if not all, placenta. However, differentiating this from organisms picked up at delivery, or during experimental handling is an on-going challenge. A recent publication examining placental tissue from healthy pregnancy concluded that these signals could not be differentiated in their cohort when rigorous contamination controls were used (Lauder et al., 2016). Furthermore, it remains impossible to assert from our data whether those organisms identified from placental samples functioned as mutualists that directly promoted the health of the developing fetus, or were commensal, and essentially neutral in their role. Metagenomic analyses, in which gene level information is retrieved, will serve to enlighten the functional role of this community in the future, in a way that 16S analyses cannot.

To summarise, these data show that sPTB is associated with the placental presence of some well-known and some novel genera of interest. Some of these sPTB cases appear to have involved infectious triggers originating from the overgrowth of a single, or small number of, organism(s), originating from the vaginal or oral cavities. This study gathered novel data from a tissue that remains largely unexplored from an unbiased microbiological perspective. The cohort consisted of a large number of spontaneous, early preterm births, for which a large number of biological replicates were also available. This improved the power of the study to detect colonisation patterns of interest. Furthermore, the use of a specifically defined 'non-spontaneous' preterm birth group was a novel addition to previous studies. These nsPTB placenta provided a comparison group that was essentially matched for GA with the sPTB cases, but very likely had different underlying aetiology. A number of interesting associations have been observed using this data. Replication of the more novel associations, such as the enrichment of the *Erwinia* genus in one of the sPTB placenta, is critical to elucidate their importance to this common and often debilitating pregnancy complication. Such follow up work will serve to further elucidate the clinical significance of these specific organisms to sPTB initiation, and potentially lead to the development of more targeted strategies to mitigate their pathogenic effect.

Chapter 7: Maternal Systemic Cytokine Profiles and Pregnancy Outcome

The experimental laboratory work in this chapter was initiated during a two-month secondment within Professor Buck's lab at Virginia Commonwealth University. This study was designed to investigate the association between maternal systemic cytokine profiles and sPTB within a subset of samples taken from the BBB cohort. The optimisation and initial experimental work were carried out personally during my visit to the United States in 2015. The remaining experimental work was completed after I left Virginia, with the generous assistance of Dr Ana Perez Miranda and Dr Anita Marinelli.

7.1 Introduction

The initiation, maintenance and completion of normal human pregnancy depend upon the successful integration of maternal endocrine and immune pathways. Traditional models of parturition, and evidence from non-human mammals, previously supported the view that endocrine changes were primarily responsible for the induction of labour (Golightly et al., 2011). However, over the past decade, evidence has been growing that refutes this endocrine-centric model of human parturition. Such evidence supports the hypothesis that in late pregnancy, it is the maternal immune system that drives the physical and biochemical changes that precede spontaneous labour in humans (Gomez-Lopez et al., 2014).

Pregnancy has often been characterized as a period of maternal immune suppression and tolerance (Sykes et al., 2012). The innate and adaptive immune systems reach a delicate balance where mothers must tolerate the semi-allogeneic fetus, whilst maintaining the capacity to fight off infection, and even reject the fetus if necessary. A switch from a T helper 1 to T helper 2 cytokine profile is thought partly to facilitate this complex demand (Sykes et al., 2012). These changes facilitate maternal acceptance of the foreign fetus, by suppressing the cell-mediated immune responses that normally detect foreign antigens. At the same time they serve to augment the antibody-mediated response to pathogens (Jamieson et al., 2006). However, as pregnancies progress into

their final weeks, this balance shifts. The culmination of normal pregnancy is labour and delivery. These events are characterised by an influx of immune cells into the uterine cavity, and a significant increase in the expression of pro-inflammatory proteins (Gomez-Lopez et al., 2014). In addition, a systemic or functional withdrawal of certain hormones, particularly progesterone, has been noted. These endocrine changes are thought to activate myometrial tissue from the quiescent state it has occupied throughout most of pregnancy (Golightly et al., 2011).

The following sections address specific evidence for the involvement of cytokines in spontaneous parturition: the main outcome of interest in the experiments reported in this chapter. Cytokines are small proteins that regulate the complex pathways of human immunity, inflammation, and haematopoesis. These readily measureable signalling molecules were the focus of the work in this part of my PhD. They have formed an important part of previous research into the role of inflammatory pathways in human labour. Systemic cytokine profiles have been investigated in a number of studies as potential biomarkers for intra-uterine infection and preterm birth risk. Cytokines are usually secreted by the cell in response to immunological stimuli. They generally act locally across both time and space. The majority of cytokines exhibit either autocrine (acting on themselves) or paracrine (acting on nearby cells) actions (Williams, 2012). However, some show endocrine activity and can function over long distances. Therefore, their presence in peripheral blood can reflect remote inflammatory events (Hebisch et al., 2004). By binding to specific membrane receptors, cytokines trigger a cascade of molecular and physical changes that can affect cell fate, function, and chemotaxis. Table 7.1 uses a common grouping structure to categorise those cytokines assayed in the proceeding experiments (Williams, 2012).

Table 7.1 – Classification of 27 cytokines used in the multiplex Bio-Rad assay used in this study by five commonly used groupings (Williams, 2012).

Cytokine family	Cytokine in assay
Interleukin	IL-1β, IL-1Ra, IL2, IL4, IL5, IL6, IL7, IL8, IL9, IL10, IL12, IL13, IL15, IL17
Interferon	IFN-γ
TNF ligand	TNF-α
Growth factor	G-CSF, GM-CSF, VEGF, FGF basic, PDGF-BB
Chemokine	CCL2, CCL3, CCL4, CCL11, CXCL10, RANTES (CCL5)

7.1.1 Parturition as an inflammatory process

Spontaneous labour is a physiological event in which the fetus, membranes, umbilical cord, and placenta are expelled from the uterus. It involves an integrated and sequential set of biochemical and physical changes, within four key tissues: the myometrium, decidua, cervix, and fetal membranes. These changes occur over a period of days or weeks and, in the majority of cases, result in a live vaginal birth. In general, changes in the structural properties of the cervix precede uterine contractility and cervical dilation. Fetal membrane rupture typically follows. The exact genetic and physiological cascades precipitating human labour are not yet fully understood.

Pro-inflammatory cytokines such as IL1, IL6, and TNF-α are secreted by peripheral leukocytes that infiltrate the uterus prior to delivery (Srikhajon et al., 2014). They are also expressed by intra-uterine cells making up decidua, amnion, and chorion tissues (Mitchell et al., 1993). Evidence for the direct involvement of these molecules in parturition comes from human, animal, and *in-vitro* work, from diverse tissues. A large body of research implicates their involvement to all stages of human labour. The role of pro-inflammatory cytokines at the culmination of both term and preterm pregnancies is particularly well supported (Tornblom et al., 2005, Hebisch et al., 2004).

7.1.1.1 Uterine contractility

Regular uterine contractions are a defining feature of the first stage of labour. They serve physically to aid the descent of the fetus through the birth canal, as well as contributing to the effacement and opening of the cervix. The process of uterine contractility itself is very well documented clinically. The molecular transition of myometrial tissue from months of muscular quiescence, to rapid and regular contractility is only partially understood.

Both term and preterm deliveries have been associated with increases in local and systemic pro-inflammatory cytokines. Samples taken from cervical secretions, and maternal blood, have implicated molecules such as IL8, IL6, IL-1β, TNF-α and CC type chemokines (e.g. CCL3, CCL4) in both events (Tency, 2014, Arntzen et al., 1997, Hebisch et al., 2004, Shobokshi and Shaarawy, 2002). Expression of CCL2, an important monocyte chemoattractant, was markedly increased in myometrial tissue from labouring as opposed to quiescent samples (Esplin et al., 2005). Cytokines such as IL1β,

IL6 and IL8, all have known pro-inflammatory activity and are present in the myometrium in both term and preterm labour (Young et al., 2002, Elliott et al., 2000). Such cytokines are thought to be involved in the transcriptional activation of a cassette of so-called 'contraction-associated proteins' (CAPs), such as oxytocin receptors, connexin-43 (cx-43), and prostaglandins. Prostaglandins are some of the most well-known initiators of uterine contractions. They are also involved in membrane degradation prior to rupture (Olson, 2003). Prostaglandins are used clinically to induce labour and cervical ripening in women. *In-vitro* work has shown that both IL1β and TNF-α stimulate prostaglandin production (Molnár et al., 1993, Leighton and Pfeilschifter, 1990). This is likely facilitated via the prolific transcription factor, NFkB (Mohan et al., 2004, Christiaens et al., 2008).

In-vitro stimulation with pro-inflammatory IL1β, enhanced Ca^{2+} presence in smooth muscle myometrial cell lines that were grown from mothers undergoing routine CS (Tribe et al., 2003). The addition of IL1β to the cell populations both increased the uptake of calcium into the cells, as well as augmenting its release from the intracellular sarcoplasmic reticulum compartment. Given the central role of Ca^{2+} ions to smooth muscle activity, this evidence provides a potential mechanistic link between pro-inflammatory cytokine presence and myometrial activity.

7.1.1.2 Cervical ripening

The softening and dilation of the cervix during labour involves extensive remodelling of the collagen in its extracellular matrix. Molecular evidence again highlights the involvement of inflammatory mechanisms and cytokines in the effacement of the cervix during labour. One of the first pieces of evidence supporting this association was the observation that leukocytes of various kinds infiltrate the cervix at both preterm and term delivery (Liggins, 1981). Immuno-active cells, such as macrophages and neutrophils, have been observed in cervical samples at concentrations 6 to 10 fold higher at delivery than during the first trimester (Bokstrom et al., 1997). Such cells both release, and are potently attracted by, pro-inflammatory cytokines a such as IL8 (Ekman-Ordeberg and Dubicke, 2012). Leukocytes and cervical smooth muscle cells also express collagenolytic enzymes such as matrix metalloproteinases 1, 3, and 9, which can directly degrade the cervical extra-cellular matrix (Watari et al., 1999).

In addition to changes in the key cellular populations in the cervix during labour, the molecular profile of the labouring cervix is strongly pro-inflammatory. Dubicke et al. (2010) observed major changes in the concentrations of both pro and anti-inflammatory mRNA and protein from labouring cervices. These changes were observed regardless of GA, indicating a common mechanism in both preterm and term cervical dilation. Compared to non-labouring tissue, mRNA concentrations of the pro-inflammatory cytokines IL1α, IL1β, IL13, and IL10, were increased in tissue from labouring cervices. These results document a suite of changes that are indicative of an association between labour, cervical ripening, and a pro-inflammatory response.

7.1.2 Cytokines and sPTB

Many of the immunological changes associated with labour are thought to be common, regardless of GA (Dubicke et al., 2010). It is therefore reasonable to hypothesise that it is in the instigation of labour, rather than the actual process itself, that any preterm specific cytokine changes may be most prominent. An untimely disruption of maternal immune tolerance, and perhaps an activation of the Th1 response, could trigger the normal labour cascade by abnormal mechanisms (Sykes et al., 2012).

A number of events known to have strong pro-inflammatory components, such as infection, stress, and obesity, are also well known risk factors for preterm birth (Seematter et al., 2004, Lyon et al., 2003). This supports the hypothesis that inflammatory pathways contribute to preterm birth risk. It would be valuable to identify any immunological signatures associated specifically with sPTB. These signatures may be the result of pro-inflammatory events, such as infection, that precede labour. By contrast, they may be the result of physiological differences that are intrinsic to the mother or fetus. At risk mothers may present with an overactive pro-inflammatory immune response, driven by an inflammatory syndrome, infection and/or genetic variation. Alternatively, differences in maternal immunological physiology could serve to increase vulnerability to ascending infection. For example, it has been shown that compromised immune states at sites such as the cervix, which are critical to protecting the uterus from ascending infection, are associated with increased sPTB incidence (Hunter et al., 2016).

The majority of studies investigating the involvement of cytokines in sPTB, including the present one, have used maternal blood samples (Menon et al., 2011). Blood

collection is an inexpensive and safe procedure to use in clinical research. Furthermore, the identification of peripheral blood biomarkers for risk profiling, has more direct clinical utility than those identified in tissues that rely on invasive procedures, such as amniocentesis, for sampling. Increased concentrations of serum IL1β have been associated with preterm compared to term labour (Torbe et al., 2007, Vitoratos et al., 2007). Investigations using peripheral cytokine profiles and sPTB have also focused on the pro-inflammatory candidates IL6 and IL8. Although there is some evidence of a positive association between such systemic cytokines and sPTB risk, results in asymptomatic women have been mixed, and associations have been generally weak or absent (Chan, 2014). Furthermore, the large variation in study designs has made meta-analyses challenging to conduct (Menon et al., 2011).

Cases of PPROM with and without infection, have been associated with increased levels of the pro-inflammatory cytokines, IL-1β, IL6, TNF-α, IFN-γ, VEGF in maternal serum and amniotic fluid (Shobokshi and Shaarawy, 2002). Fukuda et al. (2002) recruited a cohort of women with PPROM. They used samples from cordocentises to demonstrate an inverse correlation between fetal plasma IL6 concentrations and time to spontaneous labour and delivery (Fukuda et al., 2002). In this same study, correlations were observed with fetal plasma and amniotic fluid IL6 concentrations. By contrast, no correlations were observed between fetal and maternal systemic IL6, or maternal systemic IL6 and time to delivery. The authors suggest that these local inflammatory responses occurred in response to the presence of microorganisms that were identified within the uterine compartment, such as *Ureaplasma* and *Mycoplasma* spp. This was one of the first studies to demonstrate an important, and often overlooked, contribution of the fetal immune system to cases of inflammation mediated sPTB. The authors propose that in such cases, the onset of sPTL may be as much a survival mechanism for the fetus as it is for the mother. In this context, a fetal inflammatory response serves to trigger the cascade of events that will enable it to exit an environment that has become hostile. Furthermore, this study provided rare access to the amniotic and fetal compartments prior to the initiation of labour. This provided strong evidence that an up-regulation of pro-inflammatory molecules precedes, rather than follows, the clinically observed initiation of sPTB.

Differences in cytokine profiles between preterm and term pregnancies will not necessarily be detectable throughout pregnancy. A study investigating differences in 28 cytokines in the sera of mothers who delivered before 28 weeks, or at term, found no

differences at 15-18 weeks gestation (Yang et al., 2010). This implies that differences may only be observable within certain time periods, most likely closer to delivery. Identifying this critical window in which local or systemic changes in maternal physiology predispose women to preterm delivery is essential. This will facilitate the administration of potential treatments at a time when they will be effective, but prior to the point at which labour associated changes begin and become irreversible.

In other circumstances, longer-term differences in exposure and physiology may play a role in certain sPTB pathologies. In one study, early pregnancy levels of the pro-inflammatory cytokine, TNF-α, were higher in women with prior preterm delivery. These women with higher TNF-α concentrations were additionally at higher risk of a preterm birth in the current pregnancy (Curry et al., 2008). Indeed, a recent paper has called for a re-evaluation of concepts of immunology in pregnancy. It proposes that the mother's long-term, prenatal exposure to paternal antigens, and environmental pathogens, may be involved in immunological pregnancy complications (Sisti et al., 2016).

Given the significant heritability of prematurity, it is unsurprising that the identification of susceptibility alleles associated with sPTB has been an important focus for researchers. Polymorphisms in a number of genes regulating immune responses have been associated with preterm birth risk (Crider et al., 2005). Both maternal and fetal TNF-α polymorphisms have been associated with sPTB (Crider et al., 2005). Polymorphisms associated with increased expression of the pro-inflammatory IL1β gene, and those associated with reduced expression of its receptor agonist, IL-1Ra, were enriched in sPTB groups (Crider et al., 2005). A polymorphism associated with reduced expression of the pro-inflammatory cytokine IL6 was investigated in a study by Simhan et al (2003). It was present at significantly lower frequencies in women who underwent labour before 34 weeks compared to term deliveries (Simhan et al., 2003).

7.1.3 Animal models and infectious mediated PTB

Whilst evidence from a variety of sources links intra-uterine infection, pro-inflammatory pathways, and sPTB, it remains difficult to elucidate the direction of causality in this relationship. Do inflammatory signals serve to initiate labour? Or are these signals the result of the degradative changes associated with cervical remodelling, myometrial activity, and membrane rupture that lead to an inflammatory,

protective response from the mother? These questions are particularly relevant in human studies in which access to tissues of interest prior to delivery is almost entirely restricted. For this reason, animal studies have been an important experimental model for further understanding inflammation and labour. More specifically, it is important to consider how dysregulation of this axis is a causal mechanism in infectious and inflammation mediated preterm birth.

Over the past 20 years, animal models, particularly murine, have been developed to study infection, inflammation, and sPTB in more detail. Intra-uterine or systemic injections of bacterial cells or their products, such as LPS, enable investigators to observe and manipulate underlying molecular mechanisms. Studies in mice have consistently shown that LPS or bacterial cell administration results in the induction of labour, and fetal death, prior to that observed in control groups. Evidence has shown that this pathway is mediated by intra-uterine pro-inflammatory cytokines (Kaga et al., 1996, Hirsch and Muhle, 2002, Fidel et al., 1994). Similarly, the systemic administration of pro-inflammatory cytokines, such as IL1β, have been shown to induce preterm delivery in mice (Romero et al., 1991, Romero and Tartakovsky, 1992). These studies strongly support direct causal involvement of infection and maternal/fetal inflammatory responses in cases of human sPTB.

IL10 knock-out (KO) mice have been used to demonstrate the centrality of the anti-inflammatory cytokine, IL10, to susceptibility to LPS induced preterm labour and fetal loss (Robertson et al., 2006, Christiaens et al., 2008). In IL10 null mutant mice substantially elevated expression of inflammatory cytokines was recorded. Concentrations of TNF-α, IL1, IL6, and IFN-γ were higher following LPS administration in KO, compared to wildtypes (Robertson et al., 2006). The study demonstrated a direct correlation between this anti-inflammatory molecule, and the ability of the maternal immune system to respond to infection during pregnancy.

Recent RNA-seq data from three different murine models of labour have been shown to support the hypothesis that there is a unique LPS induced path to parturition, which differs from normal parturition (Migale et al., 2016). A core set of genes were shown to be up-regulated in LPS induced labour compared to term. This was also shown when the core set was compared to a progesterone-blocking preterm model. About 80% of gene changes in the LPS model were unique and included genes belonging mainly to inflammatory pathways. Animal models from the same group have also enabled

investigations of how bacterial serotypes can have different effects on pregnancy outcome (Migale et al., 2015). Such observations highlight the importance of discerning the specific underlying infections precipitating sPTB in order to better target treatment.

When the amniotic cavity of rhesus monkeys were injected with the pro-inflammatory IL1β cytokine, a concomitant increase in TNF-α, and then prostaglandin, was observed (Baggia et al., 1996). Furthermore, increases in uterine pressure, a measure used as a proxy for contractions, were documented. These changes indicate a directional link between inflammatory molecules and labour-associated chemical and physical changes. However, in these experiments frank labour, defined by cervical ripening and continued contractions, did not occur. This indicates that proteins and processes, beyond upregulation of a single inflammatory cytokine, are necessary for sustained and successful labour. Interestingly, in a related study by the same group, inoculation of the intra-amniotic cavity with group B streptococci cells, resulted in similar changes such as an increase in TNF-α, IL6, and IL1β (Gravett et al., 1994). In this experiment, contractions and cervical changes did occur, resulting in the full 'package' of labour. These studies suggest that sustained infection will likely lead to the induction of molecular components beyond IL1β and TNF-α that were used in the previous study. Furthermore, these data demonstrate a cause-and-effect relationship between bacterial infection, inflammatory responses, and preterm labour.

There is a high prevalence of fetal demise in animal models using LPS to induce preterm labour. This raises concerns over the impact of aberrant inflammation on fetal development *in-utero*. Are expectant management and the use of tocolytics to delay labour always desirable if the fetus will be exposed to continuing infectious insult and inflammatory responses for longer? Indeed, a meta-analysis provided no evidence for improved perinatal outcomes following the use of tocolysis to delay labour (Gyetvai et al., 1999). The consequences of intrauterine inflammation on fetal and neonatal cardiopulmonary, cerebral, and renal systems can be significant. However, they are beyond the scope of this thesis and are described in an excellent review by Galinsky et al. (2013).

7.1.4 Hypotheses and aims

A substantial body of work has attempted to elucidate the causal and mechanistic pathways involved in term and preterm labour over recent decades. It is now widely accepted that labour is, at least in part, an inflammatory event, and that sPTB represents an aberrant form of this process. However, reproducibly identifying those cytokines involved in pathological outcomes, which could be useful for clinical interventions and predictive testing, has been a challenge. In this study a nested case-control design was used to attempt to identify cytokines associated with sPTB, when compared to term deliveries.

The hypothesis behind this study was that in the weeks preceding delivery, maternal systemic cytokine profiles would vary between sPTB and term births. sPTB pregnancies were hypothesised to show a more consistently pro-inflammatory signal, particularly amongst cytokines known to be associated with immune responses to bacterial infection. Additionally, any cytokines specific to preterm, as opposed to term labour, would show uniform differences when compared to both term deliveries precipitated by a spontaneous event and those in which labour never occurred. The two central aims of this study were as follows:

1) Investigate the link between maternal systemic cytokine concentrations and pregnancy outcome in the run up to delivery, using BBB samples.

2) Identify cytokines associated with preterm labour that are distinct from those involved in normal term labour, and may reflect responses to intra-uterine infectious insults.

7.2 Materials and methods

7.2.1 Cohort selection

Sera from 126 pregnancies (44 sPTB, 39 spontaneous term, and 43 non-spontaneous term), collected up to 31 days before delivery, were selected from the BBB cohort for this nested case-control study (Figure 7.1). I chose to focus on those preterm births most likely to be linked to an infectious and inflammatory pathology. Therefore, the criterion for inclusion in the case group was any preterm delivery preceded by a spontaneous membrane rupture or labour event. The term control group consisted of both spontaneous and non-spontaneous births, from healthy pregnancies (no PE, IUGR, <42 weeks). In order to capture the immunological changes related to labour and delivery, I chose only samples taken within 31 days of birth. 65% of samples were taken in the final week of pregnancy. All samples available within the BBB that adhered to the criteria stated above and were available on the day of delivery were used (N=15).

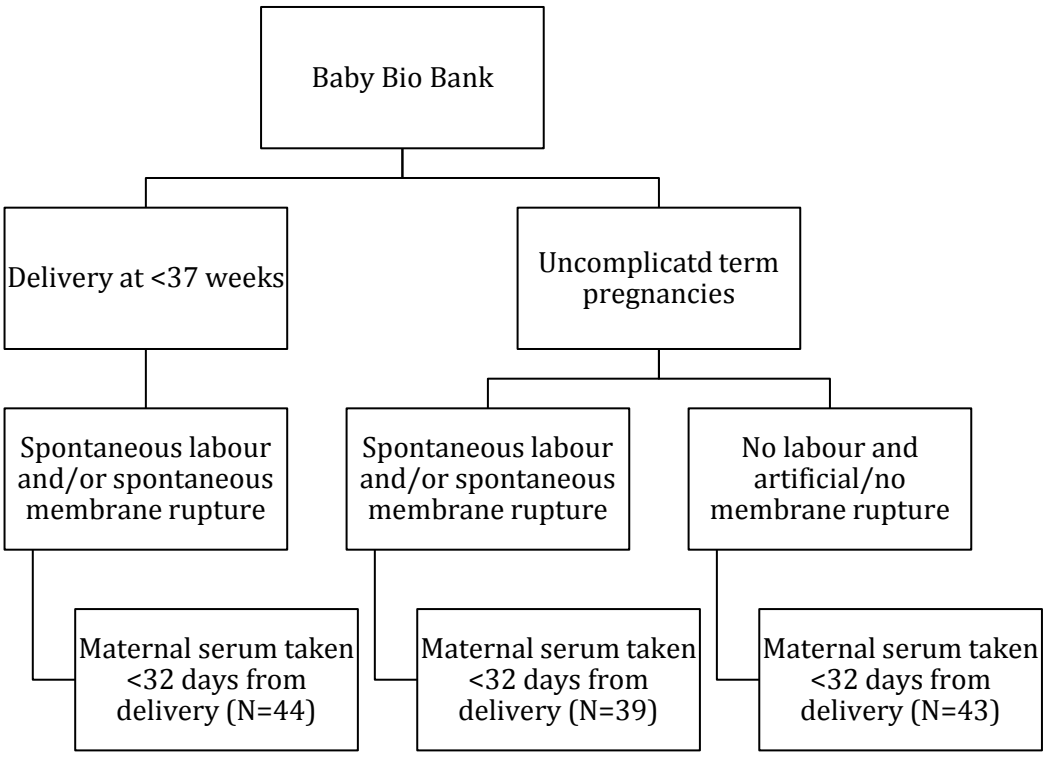


Figure 7.1 – Flow diagram documenting selection of cytokine case-control cohort

This study design facilitated the investigation of associations between cytokine profiles and pregnancy outcome. Any relationships identified could be the result of direct involvement in causal pathways, and thus hold potential for clinical developments. Alternatively, they may be the result of indirect physiological changes associated with sPTB. These could instead serve as potential late pregnancy biomarkers. Given the likelihood that the activity of some cytokines will change in the run up to delivery, samples in the three outcome groups of interest (sPTB, spontaneous term, and non-spontaneous term) were distribution matched for days before delivery at sample collection. The preterm group served as the baseline for matching, as this was the most limited collection (Figure 7.2).

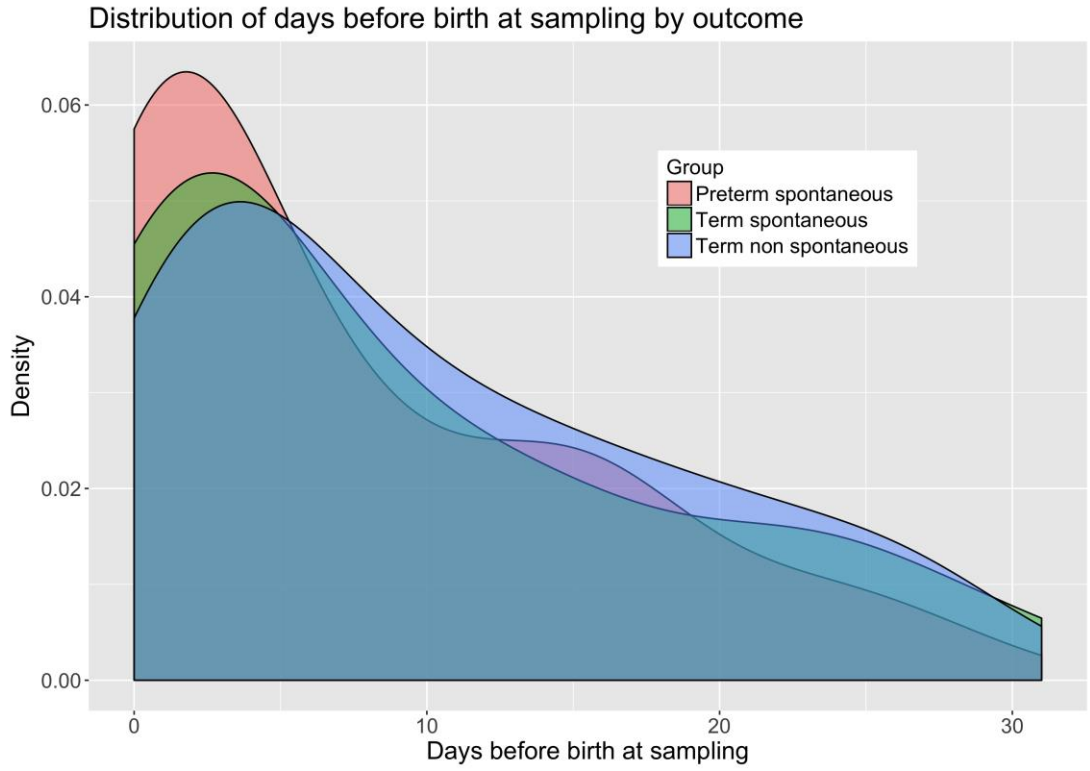


Figure 7.2 – Figure showing distribution matched profiles of days before delivery in three study groups of interest.

7.2.2 Sample collection

See section 2.4.2.

7.2.3 Multiplex ELISA assay

See section 2.4 for general methods and section 3.4 for optimisation work.

7.2.4 Statistical analyses

As discussed in section 2.5, normality of data was tested using the Shapiro-Wilcox test. Dependent on the nature of distributions, cytokine data were summarised using a number of parametric and non-parametric tests, prior to regression analyses. The lme4 package (Bates et al., 2015) was used in R to perform a linear mixed effects analysis for the prediction of cytokine concentration, from pregnancy outcome. Models were run with a random intercept term for participant number, to account for the presence of technical replicates in the sample set. An additional, independent random intercept term for batch was included if its addition improved the model fit. Otherwise batch was included as a fixed effect. In all adjusted analyses, the fixed effects of maternal ethnicity, age at booking, smoking behaviour, BMI, parity, and days before delivery at sampling were added into the model. Visual inspection of residual plots and QQ-plots confirmed that models were conforming to the assumptions of homoscedasticity and normality. P-values and degrees of freedom were estimated for the models using the Kenward-Roger approximation in the pbkrtest R package (Halekoh and Højsgaard, 2014). Samples were considered to be potentially influential outliers if their Cook's distance value was >0.1. Subsequently, such samples were removed from the dataset and models were rerun for sensitivity analyses until no further influential values were observed. Final models without any such influential values are also reported.

7.3 Results

7.3.1 Clinical dataset

Clinical data for all samples were available from the BBB clinical database. Key clinical and demographic characteristics of the cohort are summarised in Table 7.3. Diabetes data was available for less than half of the pregnancies in this group, in which only one case was observed. Therefore, this clinical category was omitted from models. 58% of deliveries were vaginal. A detailed breakdown of the labour and membrane rupture groupings in the three main clinical categories is given in Table 7.2.

Table 7.2 – Labour onset and membrane rupture method in sPTB, term spontaneous and term non-spontaneous groups for which maternal serum cytokine profiles were assessed.

Experimental grouping	Labour onset (N)		Membrane rupture method (N)	
	None	Spontaneous	None	Spontaneous
sPTB	12	32	8	36
Term spontaneous	4	35	3	36
Term non-spontaneous	43	0	43	0

Table 7.3 – Summary of maternal and pregnancy clinical data for participants of cytokine study.

Variable	Category	Sub-category	N (%)	Shapiro test (P)	Median (IQR)	Mean (SD)	Range	Total (N)	Missing	Wilcox or Fisher's exact test
GA at birth (weeks)	*All*			1.25E-9	38 (35-37)	37 (3.38)	23-41	126	0	
	Preterm			1.04E-6	35 (32.8-35)	33.4 (3.13)	23-36	44	0	W=0 P<2E-16
	Term			1.66E-5	39 (38-40)	39 (1.19)	37-41	82	0	
Birthweight (grams)	*All*			3.00E-4	3180 (2540-3450)	2980 (799.1)	540-4500	125	1	
	Preterm			0.9	2260 (1620-2540)	2150 (699.18)	540-2540	43	1	W=190.5 P=3.08E-16
	Term			7.00E-3	3310 (3150-3690)	3409.18 (419.63)	2580-4500	82	0	
Maternal age at booking	*All*			0.033	33 (29-37)	32.5 (5.77)	16-46	126	0	
	Preterm			0.44	32 (28-31.5)	31.5 (5.98)	20-42	44	0	W=1553 P=0.199
	Term			0.034	33.5 (30-33.1)	33.1 (5.62)	16-46	82	0	
GA at blood draw (weeks)	*All*			7.41E-9	7 (1-14)	8.79 (8.47)	0-31	126	0	
	Preterm			4.1E-5	3.5 (1-13)	7.39 (7.92)	0-28	44	0	W=1488 P=0.105
	Term			2.88E-6	7 (2-15.5)	9.55 (8.71)	0-31	82	0	

Maternal BMI (weight/ height^2)	*All*	Underweight	5 (3.97)			126	0
		Normal	106 (84.13)				
		Obese	15 (11.90)				
	Preterm	Underweight	3 (6.82)			44	0
		Normal	38 (86.36)				
		Obese	3 (6.82)				
	Term	Underweight	2 (2.44)			82	0
		Normal	68 (82.93)				
		Obese	12 (14.63)				
Maternal parity	*All*	0	62 (49.20)			126	0
		1	46 (36.51)				
		2+	18 (14.29)				
	Preterm	0	29 (65.91)			44	0
		1	8 (18.18)				
		2+	7 (15.91)				
	Term	0	33 (40.24)			82	0
		1	38 (46.34)				
		2+	11 (13.42)				

Maternal smoking behaviour at booking	*All*	Non-smoker	119 (95.97)				
		Smoker	5 (4.03)		124	2	
	Preterm	Non-smoker	40 (90.91)				
		Smoker	4 (9.09)		44	0	Fisher's P=0.053
	Term	Non-smoker	79 (98.75)				
		Smoker	1 (1.25)		80	2	
Maternal ethnicity	All	Asian/Asian British	9 (7.34)				
		Black/Black British	13 (10.66)				
		White/White British	77 (63.11)		122	4	
		All others	23 (18.85)				
	Preterm	Asian/Asian British	2 (4.76)				
		Black/Black British	3 (7.14)				
		White/White British	27 (64.29)		42	2	
		All others	10 (23.81)				Fisher's P=0.67
	Term	Asian/Asian British	7 (8.75)				
		Black/Black British	10 (12.50)				
		White/White British	50 (62.50)		80	2	
		All others	13 (16.25)				

7.3.2 Missingness and batch effects

This assay did not quantify all cytokines with equal efficiency, and some cytokines were more likely to have values that fell below the LOD, or outside of reliable ranges of the standard curve than others. It was important to explore the completeness of our experimental dataset, and adjust missingness appropriately, as outlined in section 3.4.2.1, before proceeding with downstream analyses.

Any fluorescence value that fell below the LOD, defined by the blank sample, was recoded as zero. A small number of cytokines accounted for the majority of zero values in the assay: IL2, IL15, CCL2, and GM-CSF. 17 observations had a high (>20%) coefficient of variation between repeat measures. These observations were deemed to be unreliable and were recoded as not available (NA).

259 observations fell below the range of the reliable standard curve. These values were recoded as half the minimum observed reliable value per plate. There was one occurrence, IL8 on plate 6, in which no observed value fell within the reliable range of the standard curve, due to a particularly poorly run curve in which only 7 out of 9 standards were within 70-130% efficiency. The assay had been reliable on all previous plates for IL8. Therefore, to minimise loss of data, the observed values, which fell outside of the reliable standard curve range, but were extrapolated by the software, were used. 44 total observations were above the reliable range of the standard curve. All of these values were from RANTES. These were recoded as the maximum observed reliable value for each plate.

526 observations, 14% of the total, were adjusted for missingness. Per cent missingness per plate was distributed across the cytokines in a non-random manner. The majority was the result of observations below the LOD of the assay (see Appendix Table A 17, page 265 for batch by batch breakdown). Missingness was focused in results from a small number of cytokines: IL2, IL15, GM-CSF, and CCL2.

Variation across batches had been approximated during optimisation (section 3.4.2.4). Correlations were high between replicates but batch effects were still clearly visible. The final two experimental batches were run with kits with a different batch of standards to the optimisation runs and the first four experimental plates. Although the manufacturers predict only minimal variation across such standard batches, it was

important to capture the nature of any differences between early and late batches. 12 experimental replicates were run between the first four and final two plates. Case and control samples were present across all plates. The analysis showed that reproducibility of individual cytokines varied considerably between these two standard sets.

The ICC values of repeats from the optimisation experiments and from the experimental replicates, were calculated and compared (Appendix Table A 18, page 266). Based on these calculations, a number of cytokines were deemed to show poor reproducibility. This was defined as either a low ICC value in both optimisation and experimental analyses (e.g. IL4), or unacceptably low reproducibility in technical replicates from experimental samples, due to the change in standard batches (e.g. IL9). On this basis, cytokines IL4, IL7, IL9, IL13, IL15, CXCL10, IFN-γ, PDGFbb, RANTES, and G-CSF were not included in final models.

For the 17 cytokines taken forward, models with and without batch as a random effect, were compared as discussed in section 7.2.4. Based on these comparisons, IL-1Ra, IL2, IL6, IL8, IL10, and CCL2 were run as models with only participant number as a random intercept, and batch as a fixed effect. IL1β, IL5, IL12, IL17A, CCL11, FGF-basic, GM-CSF, CCL3, CCL4, TNF-α, and VEGF were all run with the effects of participant identity and batch modelled as two independent random intercepts (see Appendix Table A 19, page 266 for model comparisons).

7.3.3 Summary of cytokine concentrations

The general distributions and characteristics of all 17 reliably assayed cytokines are displayed in Table 7.4. The range in concentrations varies widely by cytokine. In certain cases, zeros dominated the distribution (e.g. IL2). Furthermore, many of the cytokines were highly skewed. Therefore, for all further analyses, cytokine concentrations were log+1 transformed. This served to improve the fit of downstream regression models (data not shown).

Table 7.4 – Summary of results from cytokine experiments for 17 assayed proteins for which reproducible data were available. Concentrations in pg/ml.

Cytokine	Total (N)	Null values (N)	Missing (N)	Median (range)	Mean (SD)
IL1β	140	0	1	2.2 (0.69-33)	2.9 (3.5)
IL-1Ra	140	1	1	100 (0-3900)	180 (400)
IL2	140	80	1	0 (0-260)	8.3 (32)
IL5	140	23	0	3.8 (0-130)	6.6 (12)
IL6	140	0	3	9.1 (0.05-340)	18 (39)
IL8	140	0	0	25 (9.9-38000)	440 (3600)
IL10	140	7	0	8.8 (0-1500)	39 (150)
IL12	140	1	1	16 (0-2900)	52 (250)
IL17A	140	3	0	76 (0-430)	94 (71)
CCL11	140	0	0	44 (18-140)	50 (25)
FGF-basic	140	2	0	77 (0-300)	88 (54)
GM-CSF	140	76	2	0 (0-530)	49 (100)
CCL2	140	81	0	0 (0-5200)	81 (520)
CCL3	140	2	1	4.6 (0-53)	6.2 (6.5)
CCL4	140	0	0	110 (37-650)	120 (70)
TNF-α	140	0	0	41 (18-990)	54 (84)
VEGF	140	10	0	17 (0-1100)	38 (110)

7.3.4 Cytokine profiles and pregnancy outcome

Two categorisations of predictor variables were used when modelling the association between pregnancy outcome and the 17 cytokines that passed quality control. The first was a two-group test in which cytokine concentrations in all term samples were compared to those from sPTB deliveries. The second was a three-group model in which both term spontaneous and term non-spontaneous deliveries were compared to sPTB outcome. The latter model was used to investigate whether any observed differences between term and sPTB cytokine concentrations were a function of the non-spontaneous nature of many of the term group samples, rather than specific molecular mechanisms involved in sPTB pathology.

In unadjusted models, IL-1Ra, IL2, IL6, CCL11, GM-CSF, and CCL4 concentrations were all significantly associated with sPTB (see Appendix Table A 20 and Table A 21, pages 267-269 for full tables). Following adjustment for days before birth at sampling, maternal BMI, maternal age at booking, maternal ethnicity, maternal smoking

behaviour, batch (fixed or random), and parity, associations with IL6 disappeared, and IL8, IL10, CCL2, and IL5 associations became significant. Associations were classified into one of three groups after adjustment for confounding: (1) cytokines increased in sPTB cases; (2) cytokines decreased in sPTB cases; and (3) cytokines not associated with pregnancy outcome in this cohort. A summary of these groupings from final adjusted models is provided in Figure 7.3.

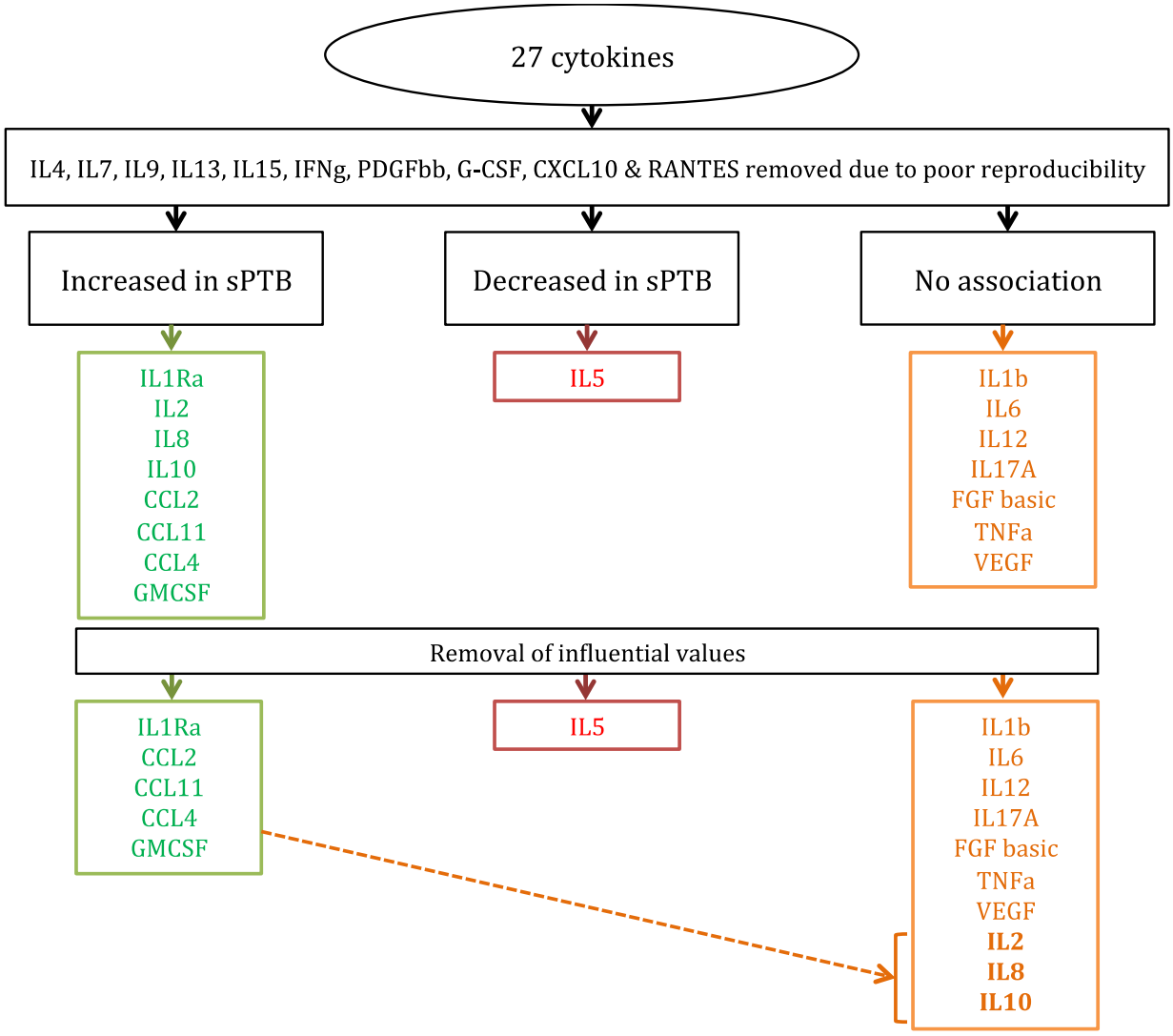


Figure 7.3 – Schematic showing those cytokines increased (green), decreased (red), and not associated with sPTB in adjusted linear regression models, before and after removal of potential influential samples. Arrow shows the three cytokines associated with sPTB that loose significance following influential sample removal.

Table 7.5 – Results from adjusted linear regression analyses quantifying difference in concentrations (pg/ml) of IL2, IL8, and IL10, between sPTB (baseline) and term or term spontaneous/non-spontaneous using whole cohort.

Cytokine	Whole cohort			
	Outcome	β (95% CI), P	Outcome	β (95% CI), P
IL2	Term	-0.63 (-1.19--0.07), 0.041	Term spontaneous	-0.62 (-1.24-0.01), 0.074
			Term non-spontaneous	-0.64 (-1.26--0.01), 0.064
IL8	Term	-0.52 (-0.91- -0.14), 0.014	Term spontaneous	-0.54 (-0.97- -0.11), 0.02
			Term non-spontaneous	-0.50 (-0.94 - -0.061), 0.039
IL10	Term	-0.55 (-1.12-0.03), 0.083	Term spontaneous	-0.71 (-1.35--0.08), 0.043
			Term non-spontaneous	-0.37 (-1.02-0.27), 0.29

Table 7.6 - Results from adjusted linear regression analyses quantifying difference in concentrations (pg/ml) of IL2, IL8, and IL10, between sPTB (baseline) and term or term spontaneous/non-spontaneous following removal of influential samples.

Cytokine	Following removal of influential samples			
	Outcome	β (95% CI), P	Outcome	β (95% CI), P
IL2	Term	-0.3 (-0.78-0.19), P=0.25	Term spontaneous	-0.22 (-0.75-0.32), P=0.45
			Term non-spontaneous	-0.38 (-0.92-0.17), P=0.20
IL8	Term	-0.073 (-0.3- 0.041), P=0.45	Term spontaneous	-0.024 (-0.19 -0.13), P=0.78
			Term non-spontaneous	-0.0065 (-0.18 0.16), P=0.94
IL10	Term	-0.46 (-0.99-0.09), P=0.11	Term spontaneous	-0.58 (-1.17-0.03), P=0.074
			Term non-spontaneous	-0.33 (-0.93-0.29), P=0.32

Cytokines IL2, IL8, and IL10 were associated with pregnancy outcome in the preliminary adjusted models. However, after models were checked for assumptions and potential influential values removed, the associations disappeared and were no longer significant (Table 7.5 and Table 7.6). Table 7.7 summarises those cytokines in which associations persisted through sensitivity analyses. See Appendix Table A 22-Table A 33, pages 271-282, for full tables of all significant models pre and post influential sample removal model.

Concentrations of IL-1Ra were significantly increased in sPTB cases, both in two and three group comparisons. In sample 336, no detectable protein was observed. This was identified using Cook's Test as an influential value, along with samples 1226 and 2388. These latter two samples came from women with sPTBs and the two highest concentrations of IL-1Ra. Following the removal of these three potential influential values, the significant association between IL-1Ra concentration and sPTB persisted in both models, but the effect size almost halved in both cases (Table 7.7).

CCL2 was also present at significantly higher concentrations in sPTB cases. No influential values were identified via diagnostic analyses. However, a large proportion of women had either no CCL2 in their sera at the time of sampling, or levels below the LOD of the assay, represented by zeros in the assay. This large proportion of zeros influenced the fit of regression analyses. Therefore, an alternative model was run to try to improve this fit by more appropriately modelling these zero values. Using an adjusted, binomial multi-level regression model, the relationship between pregnancy outcome and presence (>0 pg/ml) or absence (0 pg/ml) of CCL2 was investigated. In line with the linear model, women with detectable CCL2 protein in their serum samples had significantly lower odds of having a term birth. This relationship was significant when term births were grouped as one outcome (OR sPTB/term = 0.1 P=0.0015), or as two (OR sPTB/term spontaneous = 0.13, P=0.007, OR sPTB/non-spontaneous term = 0.081, P=0.001). The comparison of sPTB with non-spontaneous term births showed the largest difference in odds. CCL2 was associated with sPTB in this cohort in a dose dependent manner. In addition, the mere presence of CCL2 in maternal sera appears to be a biomarker for sPTB in these women.

CCL11 was higher in sera from sPTB pregnancies than term births, when considered both together and separately by spontaneous/non-spontaneous outcome. This relationship was attenuated, but persisted, after the removal of one significant outlier.

CCL4 concentration was also associated with sPTB outcome in the two-group model. However, this relationship was one of the weakest of all significant associations. When the three-group model was run, both term groups still had, on average, a reduced CCL4 concentration compared to sPTB cases. However, this effect was only significant for the comparison with non-spontaneous term births.

GM-CSF was higher in sPTB cases than term births. When investigated further using the three-group model, the relationship was only significant for spontaneous term compared to sPTB births. However, the effect sizes of both groups were similar, suggesting this may have been an issue of power. GM-CSF was also a cytokine that had a substantial number of zero values. However, when these results were split by spontaneous/non-spontaneous outcomes, no significant association was observed. A non-significant trend of lower odds of a term birth when GM-CSF was detectable was observed (data not shown). No influential values were identified for analyses involving GM-CSF.

IL5 was the only cytokine that was significantly increased in term compared to sPTB cases. However, when the three-group model was run, it was clear that the majority of this association was accounted for by differences between sPTB and non-spontaneous term births. With the removal of potentially influential samples, these relationships persisted. As this pattern was strongly indicative of an association with spontaneous events rather than outcomes by GA, this was also investigated. When IL5 concentration was compared between pregnancies culminating in a non-spontaneous outcome and a spontaneous outcome, it was significantly reduced in the spontaneous group (Beta=-0.51 95% CI=-0.84- -0.19, P=0.0022). This association supports the hypothesis that IL5 is related to pregnancy maintenance, and its reduction may be a hallmark of spontaneous parturition.

Table 7.7 - Summary table of significant associations between cytokine concentrations (pg/ml) and pregnancy outcome for two and three group models using a linear regression model adjusted for confounders.

Cytokine	Predictor (baseline =sPTB)	2 group model (β (95% CI), P)		Predictor (baseline=sPTB)	3 group model (β (95% CI) and P)	
		Adjusted model	After removal of outliers		Adjusted model	After removal of outliers
IL-1Ra	Term	-0.73 (-1.09--0.37), 0.00024	-0.38 (-0.63--0.13), 0.0054	Term spontaneous	-0.8 (-1.2--0.39), 0.00033	-0.41 (-0.69--0.14), 0.0066
				Term not spontaneous	-0.66 (-1.07--0.26), 0.0027	-0.35 (-0.62--0.07), 0.022
CCL2	Term	-1.66 (-2.47--0.86), 0.00019	-	Term spontaneous	-1.63 (-2.53--0.74), 0.0010	-
				Term not spontaneous	-1.69 (-2.59--0.8), 0.0006	-
IL5	Term	0.27 (-0.1-0.64), 0.15	0.43 (0.07-0.78), 0.019	Term spontaneous	0.01 (-0.4-0.41), 0.969	0.15 (-0.24-0.54), 0.44
				Term not spontaneous	0.52 (0.11-0.92), 0.012	0.61 (0.23-1), 0.0021
CCL11	Term	-0.32 (-0.47--0.16), 0.000056	-0.24 (-0.41--0.07), 0.0049	Term spontaneous	-0.34 (-0.51--0.15), 0.00015	-0.24 (-0.44--0.05), 0.011
				Term not spontaneous	-0.3 (-0.48--0.12), 0.0013	-0.24 (-0.43--0.05), 0.014
GM-CSF	Term	-1.06 (-2.01--0.09), 0.020	-	Term spontaneous	-1.27 (-2.33--0.19), 0.013	-
				Term not spontaneous	-0.87 (-1.91-0.21), 0.091	-
CCL4	Term	-0.18 (-0.35--0.01), 0.034	-	Term spontaneous	-0.13 (-0.32-0.06), 0.16	-0.1 (-0.29-0.09), 0.30
				Term not spontaneous	-0.23 (-0.42--0.04), 0.016	-0.22 (-0.41--0.03), 0.022

7.4 Discussion

This study has shown that the late pregnancy maternal systemic profiles of a number of cytokines are associated with sPTB. To differing degrees, IL-1Ra, CCL2, CCL4, CCL11, and GM-CSF were all up-regulated in maternal serum taken from mothers who had a sPTB, compared to those who had a term birth. By contrast, concentrations of IL5 tended to be lower among maternal serum collected from sPTB pregnancies, and in women with any spontaneous event. The remaining 21 cytokines in the assay were either omitted due to poor performance in assay optimisation, or showed no statistically significant association with outcome in this cohort.

There are two plausible interpretations for these observations. The first is that the differences between the two, or three, case groups of interest were a reflection of differences directly involved in the initiation, or the progression of parturition. In other words they are on the causal pathway. An alternative explanation is that these differences were not directly involved in the processes of interest, but instead reflected physiological responses to alternative events that precipitated these outcomes. For example, the up-regulation of the anti-inflammatory IL-1Ra cytokine in the sPTB group is unexpected if we interpret this change to reflect changes on the causal pathway to preterm labour. Instead, it is perhaps more plausible to interpret this specific association as a homeostatic mechanism evoked in response to the pro-inflammatory triggers hypothesised to underlie sPTB. Whilst reports of causal relationships are more useful for the development of clinical interventions, secondary associations are still of significant interest. Such indirect associations, provided they are reproducible, can still be powerfully utilised in the development of predictive screening programmes for women at risk of, or threatening, sPTB. From these data, it is only possible to hypothesise the nature of the observed associations based on prior knowledge. Future functional work is required to explore further these associations and their direct or indirect link to sPTB in this cohort.

IL-1Ra, CCL2, and CCL11 are the most promising candidates for cytokines involved in the molecular basis of sPTB in this cohort. These associations with sPTB, appear to be distinct from mechanisms involved in normal labour and membrane rupture. All three cytokines were consistently higher in sPTB deliveries when compared to the term group as a whole, as well as to either term group separately. The robustness of these associations makes it unlikely that the differences in cytokine concentration can be

explained by the fact that half of the term cohort was made up of pregnancies ending in non-spontaneous events. Instead, it is hypothesised that these cytokines may be directly or indirectly involved in the specific molecular mechanisms underlying sPTB. Such mechanisms may be activated following an infectious insult or some other inflammation-inducing event, such as stress. These changes could subsequently serve to increase the chances of premature activation of the normal labour cascade. Alternatively, these associations may be a consequence, rather than a cause, of the underlying molecular mechanisms of sPTB.

The picture is less clear for the two other cytokines, which were initially shown to be up-regulated in sPTB cases, compared to term births: CCL4 and GM-CSF. Whilst the direction of change is the same in both term groups, there was a loss of significance for CCL4 when compared to the term spontaneous group, and also a loss of significance with GM-CSF, when compared to term non-spontaneous. Whilst this difference may be an issue of power reduction, due to smaller samples sizes of the sub-groups, it may also be representative of true underlying physiology. The discordance implies that at least some of the observed association between sPTB and CCL4 may be on account of the type of labour and/or membrane rupture method that occurred in these groups.

The upregulation of IL-1Ra in sera from sPTB pregnancies in both regression models is a somewhat unexpected observation based on its known functional role. This cytokine is produced by macrophages, monocytes and neutrophils, and acts as the agonist for the IL1 receptor, functionally blocking the pro-inflammatory activity of IL1. IL1 has been shown across numerous independent studies to be up-regulated during spontaneous labour (Dubicke et al., 2010, Romero et al., 1992). Such observations serve to support the hypothesis that the antagonistic, and thus anti-inflammatory, effects of IL-1Ra may be protective of sPTB. The association of IL-1Ra with preterm birth is not limited to this study. Fukuda et al. (2002) observed elevated IL-1Ra expression in amniotic fluid and umbilical venous blood from PPROM cases. The authors hypothesised that this association reflected the maternal and fetal attempts to prevent overt and damaging (infectious mediated) inflammation. IL-1Ra has also been reported to be associated with labouring, compared to non-labouring, human myometrial tissue. This is again hypothesised to be involved in the resolution of labour associated inflammation (Srikhajon et al., 2014). This up-regulation of IL-1Ra, a cytokine central to homeostasis of inflammatory pathways, may be a direct response to a concomitant increase in IL1β following events such as infection. An up-regulation in sera prior to

preterm, as opposed to term delivery, may reflect the more extreme inflammatory events that precede preterm labour. This association may reflect an unsuccessful attempt by the immune system to suppress pro-inflammatory pathways and prevent sPTB, in the face of an infectious challenge. Given this observation, it is notable that we did not see any observable change in IL1β profiles in this study. It is possible that such changes were more localised to the site of infection. To our knowledge, only one other study has reported an association between peripheral IL-1Ra in maternal plasma and preterm birth (Brou et al., 2012). In a similarly sized study, women in preterm labour had a significantly higher concentration of Il1Ra than women in term labour in both maternal and fetal cord plasma.

In this study, three functionally and structurally related proteins, CCL2, CCL4, and CCL11, which are all located on chromosome 17, were up-regulated in sPTB. CCL11 and CCL2 showed more consistent associations with the sPTB phenotype. The CC-type chemokines are a class of small cytokines that act primarily as chemo-attractants within immune-modulatory pathways. Some are considered homeostatic and key to cell migration and normal tissue development and maintenance. Others are pro-inflammatory and are key to the direction of leukocytes, primarily macrophages, to sites of infection. These molecules exert their functionality through a large family of G-protein coupled receptors.

Both CCL2 and CCL4 can mediate inflammatory responses to bacterial infection (Williams, 2012). Mestan et al. (2009) showed that these two molecules, as well as CCL3, were raised in the cord blood of preterm versus term infants. By contrast, CCL2 was shown to be down regulated in cervical secretions from preterm compared to term mothers (Hunter et al., 2016). However, this observation is not necessarily inconsistent with a systemic or intra-uterine increase in this pro-inflammatory molecule. These authors hypothesise that it is precisely a lack of CCL2 at the cervices of these women, and thus a reduction in leukocyte recruiting capacity, that makes them vulnerable to infection by ascending microorganisms. By contrast, a systemic or intra-uterine increase in CCL2 may reflect response to an already present infection, as opposed to defensive mechanisms in place at the cervix. The chemo-attractant properties of these cytokines may be involved in the recruitment of leukocytes into the chorio-decidual space in those sPTBs precipitated by infectious events.

IL5 is a Th2 cytokine. These cytokines are hypothesised to predominate the maternal inflammatory profile during pregnancy (Sykes et al., 2012). IL5 is key to eosinophil activation and has long been shown to be involved in allergic diseases such as asthma. The results from the present study support the idea that IL5 may be a 'pregnancy' molecule. In both term and preterm groups in which a spontaneous event had occurred, there was a reduction in average IL5 concentration. This implies that reductions in IL5 may be part of the physiological changes that occur during, or prior to, both term and preterm spontaneous parturition.

There are a number of limitations to this study. In the first instance the scope of molecules investigated was limited by the design of a pre-defined assay. Furthermore, of those cytokines that were included, a large number were omitted from downstream analyses due to their poor performance in optimization tests. Another limitation of this study was sample size. The assay was expensive and the available samples within the timeframe outlined in the study design were limited. This may have limited the power of the study to detect smaller effect sizes. Furthermore, associations with additional outcomes of interest, for example very early preterm births or pregnancies associated with cervical suture, were not possible given the number of samples available. More complex models involving interactions of cytokines themselves were also not investigated here. It is likely that such network analyses will be important if understanding of the complex interplay of immune molecules is to be improved.

Given the nature of the BBB resource, longitudinal data was not available for individuals included within this study. As data was only available at a single time point this precluded temporal interpretation of cytokine associations with sPTB. Increased variation was also potentially introduced into the dataset by the fact that cytokines were assayed within a window of time prior to delivery rather than one static time point. Although this may have reduced the power to detect some associations, the majority of cytokines were assayed within 7 days of delivery.

Finally, it was unfortunate that there was not a larger overlap between the infection based assays of this thesis and this immunologically focused study. This would have provided an exciting opportunity to test the underlying assumption that any aberrant inflammation observed in the sPTB group was the result of an intra-uterine infectious insult. However, the samples were not available within the cohort to carry out such comparisons with any power.

Results from this study support the hypothesis that aberrant inflammation, distinct from the normal processes involved in term labour, is involved in the period preceding and during spontaneous preterm parturition. The work described in this chapter used a previously unexplored cohort to investigate how variation in concentrations of commonly occurring cytokines, in the days and hours before delivery, associate with sPTB. Most notably, there was a consistent upregulation of the three chemokines, CCL2, CCL4, and CCL11, in women who experienced a sPTB event. It is plausible that these patterns are representative of underlying aetiologies involved in preterm birth. It is reasonable to hypothesise that at least some of the observed increase in these molecules associated with leukocyte activity was a response to underlying maternal infection. This hypothesis is supported by the associations between placental infection and sPTB described in the previous chapters of this thesis, as well as evidence from independent studies linking infection with sPTB. Further work is required to determine whether the patterns discussed above are part of the causal pathway in preterm birth, or are instead biomarkers of underlying inflammatory events. Studies such as this help to elucidate which molecules are involved in inflammation and infectious mediated sPTB that may be viable candidates for future clinical developments in the field.

Chapter 8: Conclusion

The overarching objective of this thesis was to contribute to the improvement of the prediction, prevention, and management of preterm birth, and ultimately, to the health of infants born to such pregnancies. Four key objectives were set out at the start of this work. Taken together, these aimed to characterise the association between placental infection and maternal systemic inflammation, and preterm birth, in a large UK cohort. Specifically, the first aim was to collate and characterise the clinical data and biological samples held within the BBB. The outcome of this is described in Chapter 4 and is reviewed in a recent publication (Leon et al, 2016), which is available as a useful resource for future BBB based research. Next, a case control cohort of preterm and healthy term pregnancies was compiled. Samples from these pregnancies were investigated, for both quantitative and qualitative differences in the placental bacterial DNA composition of preterm and term pregnancies. In recognition of recently published, novel observations describing an apparently unique and universal 'placental microbiome', we investigated the existence of such a signature in our own placental samples. The final aim of this PhD was to examine maternal systemic cytokine profiles in the run up to delivery, with the intention of improving understanding of maternal inflammatory responses associated with preterm birth. This was conducted in a separate BBB cohort of spontaneous preterm, and term births.

The research described within this thesis, has demonstrated that certain placental microbial signatures, and maternal immunological states, associate with gestational length and pregnancy outcome. The studies that have been described were conducted on a unique and previously unexplored cohort of pregnancies, a significant number of which were born preterm with unknown aetiology. The examination of the broad range placental microbiome reported here, represents one of the largest studies to date carried out in this area. Our cohort contained a particularly large and valuable group of spontaneous, early, preterm births, taken from a multi-ethnic, UK population.

The link between infection and sPTB is well established and has been discussed at length in this thesis (Goldenberg et al., 2000). However, the specific role of the placenta in this relationship, as a site potentially harbouring both protective and pathogenic organisms during gestation, remains a topic of debate. Well-established qPCR and

protein assays, as well as state-of-the-science NGS techniques, with analytical strategies developed specifically for this cohort, were used to explore this theme further. The key findings in this thesis demonstrated that sPTB occurrence was indeed associated with both a quantitative and qualitative enrichment of certain 16S placental transcripts. These observations serve to support previous research that has reported an enrichment of bacterial DNA, and cells, in placental tissue collected from preterm pregnancies (Onderdonk et al., 2008a, Onderdonk et al., 2008b, Prince et al., 2016, Rours et al., 2011, Stout et al., 2013).

A higher 16S load was observed in placental tissue from sPTB deliveries, when compared to both nsPTB and term births, at both the maternal and fetal sides of the placenta. Not all of these associations remained significant after adjustment for confounders, although the direction of change was always the same, implying that some of this loss of significance may be attributable to issues of power. These observations imply that there may be a, largely overlooked, quantitative component to bacterial intra-uterine infection and preterm birth risk. Although these patterns were relatively robust, there was a large overlap in distributions of loads observed across all three groups: not all sPTB placenta had high 16S loads. Based on the detection of signal from negative extraction controls, it was not possible to conclusively assign all observed fluorescence from these samples to endogenous, placental bacterial DNA. This makes it difficult to use our qPCR data as evidence for a universal placental microbiome. If this entire signal was indeed 'real', and not from contaminants in reagents, the idea of a threshold quantity of bacterial DNA is a viable hypothesis for underlying sPTB aetiology in this cohort. However, an alternative hypothesis is that only the highest quantities of DNA observed in our experiments, were truly representative of bacterial infection during pregnancy. This latter scenario supports a more traditional concept of bacterial infection of the placenta as rare, largely pathogenic, and certainly not universal (Goldenberg et al., 2000). This distinction requires further investigation in future cohorts with a wider range of contamination controls available for analysis.

Qualitative summaries of the placental microbiome, in both preterm and term delivered placenta, were produced by sequencing the V5-V7 region of the 16S gene. The key observation from these experiments was that a number of extreme, single organism infections were reproducibly present within replicate placental samples, from a small number of individual sPTB cases. *Mycoplasma hominis*, *Erwinia* spp., and

Fusobacteria spp. were some of the most notable associated organisms. These individual, extreme cases likely underlie the majority of differential abundance observations identified using the Limma statistical model. This implies that it is a sub-section of extreme sPTB cases that account for differences in microbial signatures observed in this cohort, rather than a general difference in microbiome structure and composition between the two groups. This was supported by data from beta diversity analyses. There remained significant overlap between the OTUs observed within placenta from other sPTB cases, and the control groups used for comparison. sPTB is a diverse syndrome, precipitated by an array of underlying physiological and environmental states, with infection being just one of a number of important causes. This is supported by our data where only a subset of sPTB cases provided evidence for underlying, potentially causal, infection. However, it could also be argued that since only a very small section of placenta was examined from each pregnancy, a number of infectious cases may well have been missed if an infected part of the placenta was, by chance, not analysed.

The data from these experiments provide further evidence that both vaginal and oral bacteria are present in placental tissue, and may contribute to adverse outcomes. Importantly, these associations were observed independently of mode of delivery and other potential confounders. *Capnocytophaga* and *Fusobacteria* spp. are two such examples of organisms recovered from our placental tissue. These are both genera largely confined to the oral cavity that have, to differing degrees, been previously associated with adverse pregnancy outcomes (DiGiulio, 2012, Doyle et al., 2014, Lopez et al., 2010a). The observations by Aagard et al. (2014) that placental bacteria more closely resemble the oral, rather than vaginal, cavity could not be explicitly addressed here. However, it was certainly not the case that oral organisms made up the largest proportion of transcripts recovered.

In this study the most abundant genera present in placental tissue, whether term or preterm, tended to be vaginal and skin associated organisms, such as *Lactobacillus* spp. and *Streptococcus* spp.. Statistical analyses showed that abundances of these vaginal and skin organisms were strongly associated with delivery method. This demonstrated that this was an important determinant of colonisation patterns within this study. However, such genera were never exclusive to either method. This implies that not all of these vaginal and skin commensals were picked up during delivery, and that the concept of a low-level endogenous, healthy placental microbiome cannot be totally

rejected. Evidence from this data is inconclusive and must be further investigated, using cohorts with a more rigorously designed contamination protocol. Without such extra data, and further exploration of the involvement of delivery contaminants of placental samples, evidence for the universality of a placental microbiome remains mixed. Molecular evidence for such a phenomenon is, in light of these data, less conclusive than has perhaps been suggested by previous studies (Aagaard et al., 2014, Prince et al., 2016).

The work in this PhD has also demonstrated that variations in certain cytokines, present in maternal serum, correlate with pregnancy outcome. In line with previous work that has demonstrated an association between inflammatory phenotypes and sPTB, a number of cytokines, particularly those involved in the chemotaxis of leukocytes, such as CCL2 and CCL4, were up-regulated in maternal blood. However, this pattern was not exclusively observed in pro-inflammatory proteins. The upregulation of IL-1Ra, a cytokine that acts to suppress the pro-inflammatory actions of IL1, demonstrates that certain signatures will represent responses to adverse physiological events, rather than being directly on the causal pathway of preterm labour. The ability to identify associations within peripheral blood, whether variation is directly or indirectly involved in sPTB mechanisms, is encouraging for future clinical developments. Such work may in the future help to identify those women who are most likely to give birth early, and those who will carry to term. Such risk assessments have remained a significant clinical challenge within obstetrics to date.

This thesis has also identified, and presented strategies to mitigate, methodological challenges, specific to microbiome research in tissues of low biomass, such as the placenta. Interest in studying the human microbiome in health and disease is continuing to grow. Whilst the majority of early studies in this area have focused on well-established microbial communities in sites such as the vagina, gut, or oral cavity, there is increasing interest in documenting colonisation within tissues traditionally labelled as ‘sterile’ (Segal et al., 2013, Bassis et al., 2015). Such sites have often been overlooked in traditional microbiological work. Developing experimental, computational, and statistical approaches that have the sensitivity to identify low biomass communities, in mixed clinical samples, is an ongoing task. OTU picking algorithms, whilst constantly improving in efficiency and accuracy, remain imperfect. Reference databases are incomplete and of varying quality. Additionally, there remains a degree of inherent error within reads generated by NGS techniques. Although this is

relatively low, it contributes to phantom diversity and inaccurate OTU clustering, and thus potential misinterpretation of data. Nevertheless, in the context of such challenges, the work carried out in this thesis provides a proof of principle that signals of interest are discernible, from real patient populations and at low biomass sites.

Future studies investigating microbial presence in the placenta would benefit for the inclusion of a large and diverse range of contamination controls, against which experimental samples can be compared. It would be ideal to have as many rigorously collected controls, from all stages of sample collection and experimental work, as in the recent study by Lauder et al. (2016). However, this paper was based on experiments from samples of only 12 healthy pregnancies. The feasibility of carrying out such stringent contamination controls in large scale, real patient populations, such as the BBB, is limited. A balance between the requirements of sample size and control collection must be met by future studies.

Establishing causality between microbial states within the uterine cavity, systemic inflammatory markers, and adverse pregnancy outcomes, is complex. This endeavour is aided by both observational studies, such as the research outlined in this PhD, as well as functional *in-vitro*, and animal studies. Strengthening evidence for the long-standing associations between infection and PTB, serves to improve clinical options available to treat women at risk of, or threatening sPTB. The work in this thesis supports a model of infectious mediated sPTB in which specific organisms, rather than a complex 'preterm microbiome', elicit pathogenic outcomes. This implies that targeted antibiotic treatment of known pathogens could help improve PTB incidence and outcome. Identifying the specific contribution of vaginal versus oral species to these pathogenic outcomes, is also an important future aim that will further inform PTB prophylaxis, and therefore contribute to the improvement of maternal and infant health worldwide.

Bibliography

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J. & Versalovic, J. 2014. The Placenta Harbors a Unique Microbiome. *Science Translational Medicine,* 6**,** 237ra65.

Aagaard, K. M. 2014. Author response to comment on "The placenta harbors a unique microbiome". *Science Translational Medicine,* 6**,** 254lr3.

Abu-Amero, S., Thomas, A., White, S., Rogers, K., Miranda, A., Solanky, N., Leon, L. J., Demetriou, C., Ke, X., Stanier, S., Stanier, B., Costello, H., Tzehaie, S., Al-Olabi, L., Williamson, C., Johnson, M., Regan, L. & Moore, G. E. 2014. The Baby Bio Bank-A Legacy for Researchers Worldwide into Common Complications of Pregnancy. *Journal of General Practice,* 2:158

Acog 2001a. Intrauterine growth restriction - Clinical management guidelines for obstetrician–gynecologists. *International Journal of Gynaecology and Obstetrics,* 72**,** 85-96.

Acog 2001b. Intrauterine growth restriction. Clinical management guidelines for obstetrician-gynecologists. *International Journal of Gynaecology and Obstetrics,* 72**,** 85-96.

Africa, C. W. J., Nel, J. & Stemmet, M. 2014. Anaerobes and Bacterial Vaginosis in Pregnancy: Virulence Factors Contributing to Vaginal Colonisation. *Int J Environ Res Public Health,* 11**,** 6979-7000.

Aho, V. T. E., Pereira, P. a. B., Haahtela, T., Pawankar, R., Auvinen, P. & Koskinen, K. 2015. The microbiome of the human lower airways: a next generation sequencing perspective. *World Allergy Organ J,* 8.

Alfirevic, Z., Stampalija, T., Roberts, D. & Jorgensen, A. L. 2012. Cervical stitch (cerclage) for preventing preterm birth in singleton pregnancy. *Cochrane Database of Systematic Reviews*.

Allen, C. M. & Founds, S. A. 2013. Genetics and preterm birth. *J Obstet Gynecol Neonatal Nurs,* 42**,** 730-6.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

Arce, R. M., Barros, S. P., Wacker, B., Peters, B., Moss, K. & Offenbacher, S. 2009. Increased TLR4 expression in murine placentas after oral infection with periodontal pathogens. *Placenta,* 30**,** 156-62.

Arntzen, K. J., Lien, E. & Austgulen, R. 1997. Maternal serum levels of interleukin-6 and clinical characteristics of normal delivery at term. *Acta Obstet Gynecol Scand,* 76**,** 55-60.

Babyak, M. A. 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med,* 66**,** 411-21.

Baggia, S., Gravett, M. G., Witkin, S. S., Haluska, G. J. & Novy, M. J. 1996. Interleukin-Iβ Intra-Amniotic Infusion Induces TUmor Necrosis Factor-α, Prostaglandin Production, and Preterm COntractions in Pregnant Rhesus Monkeys. *Journal of the Society for Gynecologic Investigation,* 3**,** 121-126.

Barker, D. J. 1992. Fetal growth and adult disease. *Br J Obstet Gynaecol,* 99**,** 275-6.

Barker, D. J., Hales, C. N., Fall, C. H., Osmond, C., Phipps, K. & Clark, P. M. 1993. Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. *Diabetologia,* 36**,** 62-7.

Bassis, C. M., Erb-Downward, J. R., Dickson, R. P., Freeman, C. M., Schmidt, T. M., Young, V. B., Beck, J. M., Curtis, J. L. & Huffnagle, G. B. 2015. Analysis of the Upper Respiratory Tract Microbiotas as the Source of the Lung and Gastric Microbiotas in Healthy Individuals. *mBio,* 6.

Bates, D., Mächler, M., Bolker, B. & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *2015,* 67**,** 48.

Beck, S., Wojdyla, D., Say, L., Betran, A. P., Merialdi, M., Requejo, J. H., Rubens, C., Menon, R. & Van Look, P. F. 2010. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. *Bull World Health Organ,* 88**,** 31-8.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological),* 57**,** 289-300.

Bertina, R. M., Koeleman, B. P., Koster, T., Rosendaal, F. R., Dirven, R. J., De Ronde, H., Van Der Velden, P. A. & Reitsma, P. H. 1994. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature,* 369**,** 64-7.

Blencowe, H., Cousens, S., Chou, D., Oestergaard, M., Say, L., Moller, A. B., Kinney, M., Lawn, J. & The Born Too Soon Preterm Birth Action, G. 2013. Born Too Soon: The global epidemiology of 15 million preterm births. *Reprod Health,* 10**,** S2.

Bobitt, J. R. & Ledger, W. J. 1977. Unrecognized amnionitis and prematurity: a preliminary report. *J Reprod Med,* 19**,** 8-12.

Boggess, K. A., Moss, K., Madianos, P., Murtha, A. P., Beck, J. & Offenbacher, S. 2005. Fetal immune response to oral pathogens and risk of preterm birth. *American Journal of Obstetrics and Gynecology,* 193**,** 1121-1126.

Bokstrom, H., Brannstrom, M., Alexandersson, M. & Norstrom, A. 1997. Leukocyte subpopulations in the human uterine cervical stroma at early and term pregnancy. *Hum Reprod,* 12**,** 586-90.

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A. & Caporaso, J. G. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods,* 10**,** 57-9.

Botsis, D., Papagianni, V., Vitoratos, N., Makrakis, E., Aravantinos, L. & Creatsas, G. 2005. Prediction of preterm delivery by sonographic estimation of cervical length. *Biol Neonate,* 88**,** 42-5.

Boyd, H. A., Poulsen, G., Wohlfahrt, J., Murray, J. C., Feenstra, B. & Melbye, M. 2009. Maternal contributions to preterm delivery. *Am J Epidemiol,* 170**,** 1358-64.

Bray, J. R. & Curtis, J. T. 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs,* 27**,** 326-349.

Brodsky, D. & Christou, H. 2004. Current concepts in intrauterine growth restriction. *J Intensive Care Med,* 19**,** 307-19.

Brou, L., Almli, L. M., Pearce, B. D., Bhat, G., Drobek, C. O., Fortunato, S. & Menon, R. 2012. Dysregulated biomarkers induce distinct pathways in preterm birth. *Bjog,* 119**,** 458-73.

Camacho-Gonzalez, A., Spearman, P. W. & Stoll, B. J. 2013. Neonatal Infectious Diseases: Evaluation of Neonatal Sepsis. *Pediatr Clin North Am,* 60**,** 367-89.

Campbell, J. R. & Edwards, M. S. 1991. Capnocytophaga species infections in children. *Pediatr Infect Dis J,* 10**,** 944-8.

Caporaso, J. G., Bittinger, K., Bushman, F. D., Desantis, T. Z., Andersen, G. L. & Knight, R. 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics,* 26**,** 266-7.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., Mcdonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Tumbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. 2010b. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods,* 7**,** 335-336.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G. & Knight, R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J,* 6**,** 1621-1624.

Carr-Hill, R. A. & Hall, M. H. 1985. The repetition of spontaneous preterm labour. *Br J Obstet Gynaecol,* 92**,** 921-8.

Chambrone, L., Pannuti, C. M., Guglielmetti, M. R. & Chambrone, L. A. 2011. Evidence grade associating periodontitis with preterm birth and/or low birth weight: II: a systematic review of randomized trials evaluating the effects of periodontal treatment. *J Clin Periodontol,* 38**,** 902-14.

Chan, G. J., Lee, A. C., Baqui, A. H., Tan, J. & Black, R. E. 2015. Prevalence of early-onset neonatal infection among newborns of mothers with bacterial infection or colonization: a systematic review and meta-analysis. *BMC Infectious Diseases,* 15**,** 118.

Chan, R. L. 2014. Biochemical Markers of Spontaneous Preterm Birth in Asymptomatic Women. *BioMed Research International,* 2014**,** 8.

Charbonneau, M. R., Blanton, L. V., Digiulio, D. B., Relman, D. A., Lebrilla, C. B., Mills, D. A. & Gordon, J. I. 2016. A microbial perspective of human developmental biology. *Nature,* 535**,** 48-55.

Christiaens, I., Zaragoza, D. B., Guilbert, L., Robertson, S. A., Mitchell, B. F. & Olson, D. M. 2008. Inflammatory processes in preterm and term parturition. *Journal of Reproductive Immunology,* 79**,** 50-57.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., Mcgarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R. & Tiedje, J. M. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res,* 42**,** D633-42.

Combs, C. A., Gravett, M., Garite, T. J., Hickok, D. E., Lapidus, J., Porreco, R., Rael, J., Grove, T., Morgan, T. K., Clewell, W., Miller, H., Luthy, D., Pereira, L., Nageotte, M., Robilio, P. A., Fortunato, S., Simhan, H., Baxter, J. K., Amon, E., Franco, A., Trofatter, K. & Heyborne, K. 2014. Amniotic fluid infection,

inflammation, and colonization in preterm labor with intact membranes. *Am J Obstet Gynecol,* 210**,** 125.e1-125.e15.

Conde-Agudelo, A., Rosas-Bermúdez, A. & Kafury-Goeta, A. C. 2006. Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. *JAMA,* 295**,** 1809-23.

Crider, K. S., Whitehead, N. & Buus, R. M. 2005. Genetic variation associated with preterm birth: a HuGE review. *Genet Med,* 7**,** 593-604.

Cuiv, P. O., De Carcer, D. A., Jones, M., Klaassens, E. S., Worthley, D. L., Whitehall, V. L. J., Kang, S., Mcsweeney, C. S., Leggett, B. A. & Morrison, M. 2011. The Effects from DNA Extraction Methods on the Evaluation of Microbial Diversity Associated with Human Colonic Tissue. *Microbial Ecology,* 61**,** 353-362.

Curry, A. E., Vogel, I., Skogstrand, K., Drews, C., Schendel, D. E., Flanders, W. D., Hougaard, D. M. & Thorsen, P. 2008. Maternal plasma cytokines in early- and mid-gestation of normal human pregnancy and their association with maternal factors. *Journal of Reproductive Immunology,* 77**,** 152-160.

Da Fonseca, E. B., Bittar, R. E., Carvalho, M. H. & Zugaib, M. 2003. Prophylactic administration of progesterone by vaginal suppository to reduce the incidence of spontaneous preterm birth in women at increased risk: a randomized placebo-controlled double-blind study. *Am J Obstet Gynecol,* 188**,** 419-24.

De Andrade Ramos, B., Kanninen, T. T., Sisti, G. & Witkin, S. S. 2014. Microorganisms in the Female Genital Tract during Pregnancy: Tolerance versus Pathogenesis. *Am J Reprod Immunol*;73:383-389

De Beaudrap, P., Turyakira, E., White, L. J., Nabasumba, C., Tumwebaze, B., Muehlenbachs, A., Guerin, P. J., Boum, Y., Mcgready, R. & Piola, P. 2013. Impact of malaria during pregnancy on pregnancy outcomes in a Ugandan prospective cohort with intensive malaria screening and prompt treatment. *Malar J,* 12**,** 139.

De Boer, R., Peters, R., Gierveld, S., Schuurman, T., Kooistra-Smid, M. & Savelkoul, P. 2010. Improved detection of microbial DNA after bead-beating before DNA isolation. *J Microbiol Methods,* 80**,** 209-11.

De Jager, W., Bourcier, K., Rijkers, G. T., Prakken, B. J. & Seyfert-Margolis, V. 2009. Prerequisites for cytokine measurements in clinical trials with multiplex immunoassays. *BMC Immunol,* 10**,** 52.

Delorme-Axford, E., Sadovsky, Y. & Coyne, C. B. 2014. The Placenta as a Barrier to Viral Infections. *Annu Rev Virol,* 1**,** 133-46.

Demetriou, C., Abu-Amero, S., Thomas, A. C., Ishida, M., Aggarwal, R., Al-Olabi, L., Leon, L. J., Stafford, J. L., Syngelaki, A., Peebles, D., Nicolaides, K. H., Regan, L., Stanier, P. & Moore, G. E. 2014. Paternally expressed, imprinted insulin-like growth factor-2 in chorionic villi correlates significantly with birth weight. *PLoS One,* 9**,** e85454.

Desantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G. L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol,* 72**,** 5069-72.

Digiulio, D. B. 2012. Diversity of microbes in amniotic fluid. *Semin Fetal Neonatal Med,* 17**,** 2-11.

Digiulio, D. B., Romero, R., Amogan, H. P., Kusanovic, J. P., Bik, E. M., Gotsch, F., Kim, C. J., Erez, O., Edwin, S. & Relman, D. A. 2008. Microbial prevalence, diversity and abundance in amniotic fluid during preterm labor: a molecular and culture-based investigation. *PLoS One,* 3**,** e3056.

Digiulio, D. B., Romero, R., Kusanovic, J. P., Gomez, R., Kim, C. J., Seok, K. S., Gotsch, F., Mazaki-Tovi, S., Vaisbuch, E., Sanders, K., Bik, E. M., Chaiworapongsa, T., Oyarzun, E. & Relman, D. A. 2010. Prevalence and diversity of microbes in the amniotic fluid, the fetal inflammatory response, and pregnancy outcome in women with preterm pre-labor rupture of membranes. *Am J Reprod Immunol,* 64**,** 38-57.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science,* 14**,** 927-930.

Dodd, J. M., Jones, L., Flenady, V., Cincotta, R. & Crowther, C. A. 2013. Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth. *Cochrane Database of Systematic Reviews*.

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G. & Fierer, N. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A,* 107.

Douvier, S., Neuwirth, C., Filipuzzi, L. & Kisterman, J.-P. 1999. Chorioamnionitis with intact membranes caused by Capnocytophaga sputigena. *European Journal of Obstetrics & Gynecology and Reproductive Biology,* 83**,** 109-112.

Doyle, R. M., Alber, D. G., Jones, H. E., Harris, K., Fitzgerald, F., Peebles, D. & Klein, N. 2014. Term and preterm labour are associated with distinct microbial community structures in placental membranes which are independent of mode of delivery. *Placenta,* 35**,** 1099-1101.

Dubicke, A., Fransson, E., Centini, G., Andersson, E., Bystrom, B., Malmstrom, A., Petraglia, F., Sverremark-Ekstrom, E. & Ekman-Ordeberg, G. 2010. Pro-inflammatory and anti-inflammatory cytokines in human preterm and term cervical ripening. *J Reprod Immunol,* 84**,** 176-85.

Duley, L. 2009. The global impact of pre-eclampsia and eclampsia. *Seminars in Perinatology,* 33**,** 130-7.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics,* 26**,** 2460-2461.

Edgar, R. C. & Flyvbjerg, H. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics,* 31**,** 3476-3482.

Ekman-Ordeberg, G. & Dubicke, A. 2012. Preterm Cervical Ripening in humans. *Facts Views Vis Obgyn,* 4**,** 245-53.

Elliott, C. L., Slater, D. M., Dennes, W., Poston, L. & Bennett, P. R. 2000. Interleukin 8 expression in human myometrium: changes in relation to labor onset and with gestational age. *Am J Reprod Immunol,* 43**,** 272-7.

Elovitz, M. A. & Mrinalini, C. 2004. Animal models of preterm birth. *Trends Endocrinol Metab,* 15**,** 479-87.

Elovitz, M. A., Wang, Z., Chien, E. K., Rychlik, D. F. & Phillippe, M. 2003. A new model for inflammation-induced preterm birth: the role of platelet-activating factor and Toll-like receptor-4. *Am J Pathol,* 163**,** 2103-11.

Ernest, J. M. & Wasilauskas, B. 1985. Capnocytophaga in the amniotic fluid of a woman in preterm labor with intact membranes. *American Journal of Obstetrics and Gynecology,* 153**,** 648-649.

Esplin, M. S., Peltier, M. R., Hamblin, S., Smith, S., Fausett, M. B., Dildy, G. A., Branch, D. W., Silver, R. M. & Adashi, E. Y. 2005. Monocyte chemotactic protein-1 expression is increased in human gestational tissues during term and preterm labor. *Placenta,* 26**,** 661-671.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res,* 8**,** 175-85.

Fardini, Y., Wang, X., Temoin, S., Nithianantham, S., Lee, D., Shoham, M. & Han, Y. W. 2011. Fusobacterium nucleatum adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity. *Mol Microbiol,* 82**,** 1468-80.

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A. & Pradhan, S. 2013. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One,* 8**,** e76096.

Fichorova, R. N., Onderdonk, A. B., Yamamoto, H., Delaney, M. L., Dubois, A. M., Allred, E. & Leviton, A. 2011. Maternal Microbe-Specific Modulation of Inflammatory Response in Extremely Low-Gestational-Age Newborns. *mBio,* 2.

Fidel, P. L., Jr., Romero, R., Wolf, N., Cutright, J., Ramirez, M., Araneda, H. & Cotton, D. B. 1994. Systemic and local cytokine profiles in endotoxin-induced preterm parturition in mice. *Am J Obstet Gynecol,* 170**,** 1467-75.

Forsen, T., Eriksson, J., Tuomilehto, J., Reunanen, A., Osmond, C. & Barker, D. 2000. The fetal and childhood growth of persons who develop type 2 diabetes. *Annals of Internal Medicine,* 133**,** 176-82.

Fortner, K. B., Grotegut, C. A., Ransom, C. E., Bentley, R. C., Feng, L., Lan, L., Heine, R. P., Seed, P. C. & Murtha, A. P. 2014. Bacteria localization and chorion thinning among preterm premature rupture of membranes. *PLoS One,* 9**,** e83338.

Fukuda, H., Masuzaki, H. & Ishimaru, T. 2002. Interleukin-6 and interleukin-1 receptor antagonist in amniotic fluid and cord blood in patients with pre-term, premature rupture of the membranes. *Int J Gynaecol Obstet,* 77**,** 123-9.

Galinsky, R., Polglase, G. R., Hooper, S. B., Black, M. J. & Moss, T. J. M. 2013. The Consequences of Chorioamnionitis: Preterm Birth and Effects on Development. *J Pregnancy,* 2013.

Gardella, C., Riley, D. E., Hitti, J., Agnew, K., Krieger, J. N. & Eschenbach, D. 2004. Identification and sequencing of bacterial rDNAs in culture-negative amniotic fluid from women in premature labor. *Am J Perinatol,* 21**,** 319-23.

Giorlandino, C., Calugi, G., Iaconianni, L., Santoro, M. L. & Lippa, A. 1998. Spermatozoa with chromosomal abnormalities may result in a higher rate of recurrent abortion. *Fertility and Sterility,* 70**,** 576-7.

Girsen, A. I., Mayo, J. A., Carmichael, S. L., Phibbs, C. S., Shachar, B. Z., Stevenson, D. K., Lyell, D. J., Shaw, G. M. & Gould, J. B. 2016. Women's prepregnancy

underweight as a risk factor for preterm birth: a retrospective study. *Bjog*.

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B. & Chiodini, R. J. 2016. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathogens,* 8**,** 24.

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., Jorden, J. R. & Chiodini, R. J. 2015. Changes in 16s RNA Gene Microbial Community Profiling by Concentration of Prokaryotic DNA. *Journal of Microbiological Methods,* 119**,** 239-242.

Glueck, C. J., Phillips, H., Cameron, D., Wang, P., Fontaine, R. N., Moore, S. K., Sieve-Smith, L. & Tracy, T. 2000. The 4G/4G polymorphism of the hypofibrinolytic plasminogen activator inhibitor type 1 gene: an independent risk factor for serious pregnancy complications. *Metabolism,* 49**,** 845-52.

Goepfert, A. R., Jeffcoat, M. K., Andrews, W. W., Faye-Petersen, O., Cliver, S. P., Goldenberg, R. L. & Hauth, J. C. 2004. Periodontal disease and upper genital tract inflammation in early spontaneous preterm birth. *Obstet Gynecol,* 104**,** 777-83.

Goldenberg, R. L., Cliver, S. P., Mulvihill, F. X., Hickey, C. A., Hoffman, H. J., Klerman, L. V. & Johnson, M. J. 1996. Medical, psychosocial, and behavioral risk factors do not explain the increased risk for low birth weight among black women. *Am J Obstet Gynecol,* 175**,** 1317-24.

Goldenberg, R. L., Culhane, J. F., Iams, J. D. & Romero, R. 2008. Epidemiology and causes of preterm birth. *Lancet,* 371**,** 75-84.

Goldenberg, R. L., Hauth, J. C. & Andrews, W. W. 2000. Intrauterine infection and preterm delivery. *N Engl J Med,* 342**,** 1500-7.

Goldenberg, R. L., Iams, J. D., Mercer, B. M., Meis, P. J., Moawad, A. H., Copper, R. L., Das, A., Thom, E., Johnson, F., Mcnellis, D., Miodovnik, M., Van Dorsten, J. P., Caritis, S. N., Thurnau, G. R. & Bottoms, S. F. 1998. The preterm prediction study: the value of new vs standard risk factors in predicting early and all spontaneous preterm births. NICHD MFMU Network. *American Journal of Public Health,* 88**,** 233-238.

Golightly, E., Jabbour, H. N. & Norman, J. E. 2011. Endocrine immune interactions in human parturition. *Molecular and Cellular Endocrinology,* 335**,** 52-59.

Gomez-Lopez, N., Stlouis, D., Lehr, M. A., Sanchez-Rodriguez, E. N. & Arenas-Hernandez, M. 2014. Immune cells in term and preterm labor. *Cell Mol Immunol,* 11**,** 571-81.

Goncalves, L. F., Chaiworapongsa, T. & Romero, R. 2002. Intrauterine infection and prematurity. *Mental Retardation and Developmental Disabilities Research Reviews,* 8**,** 3-13.

Gosalbes, M. J., Llop, S., Valles, Y., Moya, A., Ballester, F. & Francino, M. P. 2013. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clin Exp Allergy,* 43**,** 198-211.

Gravett, M. G., Hummel, D., Eschenbach, D. A. & Holmes, K. K. 1986. Preterm labor associated with subclinical amniotic fluid infection and with bacterial vaginosis. *Obstet Gynecol,* 67.

Gravett, M. G., Witkin, S. S., Haluska, G. J., Edwards, J. L., Cook, M. J. & Novy, M. J. 1994. An experimental model for intraamniotic infection and preterm labor in rhesus monkeys. *Am J Obstet Gynecol,* 171**,** 1660-7.

Grimbizis, G. F., Camus, M., Tarlatzis, B. C., Bontis, J. N. & Devroey, P. 2001. Clinical implications of uterine malformations and hysteroscopic treatment results. *Human Reproduction Update,* 7**,** 161-74.

Grimes-Dennis, J. & Berghella, V. 2007. Cervical length and prediction of preterm delivery. *Curr Opin Obstet Gynecol,* 19**,** 191-5.

Guinn, D. A., Goepfert, A. R., Owen, J., Brumfield, C. & Hauth, J. C. 1997. Management options in women with preterm uterine contractions: a randomized clinical trial. *Am J Obstet Gynecol,* 177**,** 814-8.

Guzick, D. S. & Winn, K. 1985. The association of chorioamnionitis with preterm delivery. *Obstet Gynecol,* 65**,** 11-6.

Guzman, E. R., Shen-Schwarz, S., Benito, C., Vintzileos, A. M., Lake, M. & Lai, Y. L. 1999. The relationship between placental histology and cervical ultrasonography in women at risk for pregnancy loss and spontaneous preterm birth. *Am J Obstet Gynecol,* 181**,** 793-7.

Gyetvai, K., Hannah, M. E., Hodnett, E. D. & Ohlsson, A. 1999. Tocolytics for preterm labor: a systematic review. *Obstet Gynecol,* 94**,** 869-77.

Halekoh, U. & Højsgaard, S. 2014. A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models – The R Package pbkrtest. *2014,* 59**,** 32.

Han, Y. W., Fardini, Y., Chen, C., Iacampo, K. G., Peraino, V. A., Shamonki, J. M. & Redline, R. W. 2010. Term stillbirth caused by oral Fusobacterium nucleatum. *Obstet Gynecol,* 115**,** 442-5.

Han, Y. W., Ikegami, A., Bissada, N. F., Herbst, M., Redline, R. W. & Ashmead, G. G. 2006. Transmission of an Uncultivated Bergeyella Strain from the Oral Cavity to Amniotic Fluid in a Case of Preterm Birth. *Journal of Clinical Microbiology,* 44**,** 1475-1483.

Han, Y. W., Redline, R. W., Li, M., Yin, L., Hill, G. B. & Mccormick, T. S. 2004. Fusobacterium nucleatum Induces Premature and Term Stillbirths in Pregnant Mice: Implication of Oral Bacteria in Preterm Birth. *Infect Immun,* 72**,** 2272-9.

Han, Y. W., Shen, T., Chung, P., Buhimschi, I. A. & Buhimschi, C. S. 2009. Uncultivated Bacteria as Etiologic Agents of Intra-Amniotic Inflammation Leading to Preterm Birth. *Journal of Clinical Microbiology,* 47**,** 38-47.

Han, Y. W. & Wang, X. 2013. Mobile Microbiome: Oral Bacteria in Extra-oral Infections and Inflammation. *Journal of Dental Research,* 92**,** 485-491.

Hansen, R., Scott, K. P., Khan, S., Martin, J. C., Berry, S. H., Stevenson, M., Okpapi, A., Munro, M. J. & Hold, G. L. 2015. First-Pass Meconium Samples from Healthy Term Vaginally-Delivered Neonates: An Analysis of the Microbiota. *PLoS ONE,* 10**,** e0133320.

Harding, R. & Bocking, A. D. 2001. *Fetal Growth and Development,* Cambridge, Cambridge University Press.

Harris, J. B., H 1927. Bacterial content of the uterus at Cesarean section. *American Journal of Obstetrics & Gynecology,* 13.133:143

Harris, K. A. & Hartley, J. C. 2003. Development of broad-range 16S rDNA PCR for use in the routine diagnostic clinical microbiology service. *J Med Microbiol,* 52**,** 685-91.

Hassan, S., Romero, R., Hendler, I., Gomez, R., Khalek, N. & Espinoza, J. 2006. A sonographic short cervix as the only clinical manifestation of intra-amniotic infection. *J Perinat Med,* 34.

Hatanaka, A. R., Mattar, R., Kawanami, T. E., França, M. S., Rolo, L. C., Nomura, R. M., Araujo Júnior, E., Nardozza, L. M. & Moron, A. F. 2014. Amniotic fluid "sludge" is an independent risk factor for preterm delivery. *J Matern Fetal Neonatal Med***,** 1-6.

Hebisch, G., Neumaier-Wagner, P. M., Huch, R. & Von Mandach, U. 2004. Maternal serum interleukin-1 beta, -6 and -8 levels and potential determinants in pregnancy and peripartum. *J Perinat Med,* 32**,** 475-80.

Hill, G. B. 1998. Preterm Birth: Associations With Genital and Possibly Oral Microflora. *Annals of Periodontology,* 3**,** 222-232.

Hillier, S. L., Martius, J., Krohn, M., Kiviat, N., Holmes, K. K. & Eschenbach, D. A. 1988. A case-control study of chorioamnionic infection and histologic chorioamnionitis in prematurity. *N Engl J Med,* 319**,** 972-8.

Hillier, S. L., Nugent, R. P., Eschenbach, D. A., Krohn, M. A., Gibbs, R. S. & Martin, D. H. 1995. Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. *The Vaginal Infections and Prematurity Study Group. N Engl J Med,* 333.

Hillman, S. L., Finer, S., Smart, M. C., Mathews, C., Lowe, R., Rakyan, V. K., Hitman, G. A. & Williams, D. J. 2015. Novel DNA methylation profiles associated with key gene regulation and transcription pathways in blood and placenta of growth-restricted neonates. *Epigenetics,* 10**,** 50-61.

Hirsch, E. & Muhle, R. 2002. Intrauterine Bacterial Inoculation Induces Labor in the Mouse by Mechanisms Other than Progesterone Withdrawal. *Biology of Reproduction,* 67**,** 1337-1341.

Hitti, J., Riley, D. E., Krohn, M. A., Hillier, S. L., Agnew, K. J., Krieger, J. N. & Eschenbach, D. A. 1997. Broad-spectrum bacterial rDNA polymerase chain reaction assay for detecting amniotic fluid infection among women in premature labor. *Clin Infect Dis,* 24**,** 1228-32.

Huang, B., Fettweis, J. M., Brooks, J. P., Jefferson, K. K. & Buck, G. A. 2014. The changing landscape of the vaginal microbiome. *Clin Lab Med,* 34**,** 747-61.

Hugenholtz, P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biology,* 3**,** reviews0003.1-reviews0003.8.

Hunter, P. J., Sheikh, S., David, A. L., Peebles, D. M. & Klein, N. 2016. Cervical leukocytes and spontaneous preterm birth. *Journal of Reproductive Immunology,* 113**,** 42-49.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., Fitzgerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., Versalovic, J., Wollam, A. M., Worley, K. C., Wortman, J. R., Young, S. K., Zeng, Q. D., Aagaard, K. M., Abolude, O. O., Allen-Vercoe, E., Alm, E. J., Alvarado, L., Andersen, G. L., Anderson, S., Appelbaum, E., Arachchi, H. M., Armitage, G., Arze, C. A., Ayvaz, T., Baker, C. C., Begg, L., Belachew, T., Bhonagiri, V., Bihan, M., Blaser, M. J., Bloom, T., Bonazzi, V., Brooks, J. P., Buck, G. A., Buhay, C. J.,

Busam, D. A., Campbell, J. L., Canon, S. R., Cantarel, B. L., Chain, P. S. G., Chen, I. M. A., Chen, L., Chhibba, S., Chu, K., Ciulla, D. M., Clemente, J. C., Clifton, S. W., Conlan, S., Crabtree, J., Cutting, M. A., Davidovics, N. J., Davis, C. C., Desantis, T. Z., Deal, C., Delehaunty, K. D., Dewhirst, F. E., Deych, E., Ding, Y., Dooling, D. J., Dugan, S. P., Dunne, W. M., Durkin, A. S., Edgar, R. C., Erlich, R. L., Farmer, C. N., Farrell, R. M., Faust, K., Feldgarden, M., Felix, V. M., Fisher, S., Fodor, A. A., Forney, L. J., Foster, L., Di Francesco, V., Friedman, J., Friedrich, D. C., Fronick, C. C., Fulton, L. L., Gao, H. Y., Garcia, N., Giannoukos, G., Giblin, C., Giovanni, M. Y., Goldberg, J. M., Goll, J., Gonzalez, A., Griggs, A., et al. 2012. Structure, function and diversity of the healthy human microbiome. *Nature,* 486**,** 207-214.

Hutzal, C. E., Boyle, E. M., Kenyon, S. L., Nash, J. V., Winsor, S., Taylor, D. J. & Kirpalani, H. 2008. Use of antibiotics for the treatment of preterm parturition and prevention of neonatal morbidity: a metaanalysis. *Am J Obstet Gynecol,* 199**,** 620.e1-8.

Hyman, R. W., Fukushima, M., Jiang, H., Fung, E., Rand, L., Johnson, B., Vo, K. C., Caughey, A. B., Hilton, J. F., Davis, R. W. & Giudice, L. C. 2014. Diversity of the Vaginal Microbiome Correlates With Preterm Birth. *Reproductive Sciences,* 21**,** 32-40.

Iams, J. D., Goldenberg, R. L., Meis, P. J., Mercer, B. M., Moawad, A., Das, A., Thom, E., Mcnellis, D., Copper, R. L., Johnson, F. & Roberts, J. M. 1996. The length of the cervix and the risk of spontaneous premature delivery. National Institute of Child Health and Human Development Maternal Fetal Medicine Unit Network. *N Engl J Med,* 334**,** 567-72.

Iams, J. D., Goldenberg, R. L., Mercer, B. M., Moawad, A., Thom, E., Meis, P. J., Mcnellis, D., Caritis, S. N., Miodovnik, M., Menard, M. K., Thurnau, G. R., Bottoms, S. F. & Roberts, J. M. 1998. The Preterm Prediction Study: Recurrence risk of spontaneous preterm birth. *American Journal of Obstetrics and Gynecology,* 178**,** 1035-1040.

Iams, J. D., Romero, R., Culhane, J. F. & Goldenberg, R. L. 2008. Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. *Lancet,* 371**,** 164-75.

Ikegami, A., Chung, P. & Han, Y. W. 2009. Complementation of the fadA Mutation in Fusobacterium nucleatum Demonstrates that the Surface-Exposed Adhesin Promotes Cellular Invasion and Placental Colonization. *Infection and Immunity,* 77**,** 3075-3079.

Jalava, J., Mantymaa, M. L., Ekblad, U., Toivanen, P., Skurnik, M., Lassila, O. & Alanen, A. 1996. Bacterial 16S rDNA polymerase chain reaction in the detection of intra-amniotic infection. *Br J Obstet Gynaecol,* 103**,** 664-9.

Jamieson, D. J., Theiler, R. N. & Rasmussen, S. A. 2006. Emerging infections and pregnancy. *Emerg Infect Dis,* 12**,** 1638-43.

Jaric, M., Segal, J., Silva-Herzog, E., Schneper, L., Mathee, K. & Narasimhan, G. 2013. Better primer design for metagenomics applications by increasing taxonomic distinguishability. *BMC Proceedings,* 7**,** S4.

Jeffcoat, M. K., Geurs, N. C., Reddy, M. S., Cliver, S. P., Goldenberg, R. L. & Hauth, J. C. 2001. Periodontal infection and preterm birth: results of a prospective study. *J Am Dent Assoc,* 132**,** 875-80.

Jervis-Bardy, J., Leong, L. E. X., Marri, S., Smith, R. J., Choo, J. M., Smith-Vaughan, H. C., Nosworthy, E., Morris, P. S., O'leary, S., Rogers, G. B. & Marsh, R. L.

2015. Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. *Microbiome,* 3**,** 1-11.

Johansson, M. A., Mier-Y-Teran-Romero, L., Reefhuis, J., Gilboa, S. M. & Hills, S. L. 2016. Zika and the Risk of Microcephaly. *New England Journal of Medicine,* 375**,** 1-4.

Jones, H. E., Harris, K. A., Azizia, M., Bank, L., Carpenter, B., Hartley, J. C., Klein, N. & Peebles, D. 2009. Differing prevalence and diversity of bacterial species in fetal membranes from very preterm and term labor. *PLoS One,* 4**,** e8205.

Kaga, N., Katsuki, Y., Obata, M. & Shibutani, Y. 1996. Repeated administration of low-dose lipopolysaccharide induces preterm delivery in mice: a model for human preterm parturition and for assessment of the therapeutic ability of drugs against preterm delivery. *Am J Obstet Gynecol,* 174**,** 754-9.

Kennedy, K., Hall, M. W., Lynch, M. D. J., Moreno-Hagelsieb, G. & Neufeld, J. D. 2014a. Evaluating Bias of Illumina-Based Bacterial 16S rRNA Gene Profiles. *Applied and Environmental Microbiology,* 80**,** 5717-5722.

Kennedy, N. A., Walker, A. W., Berry, S. H., Duncan, S. H., Farquarson, F. M., Louis, P., Thomson, J. M., Other Members Not Named within the Manuscript Author, L., Satsangi, J., Flint, H. J., Parkhill, J., Lees, C. W. & Hold, G. L. 2014b. The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing. *PLoS One,* 9**,** e88982.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., Haussler & David 2002. The Human Genome Browser at UCSC. *Genome Research,* 12**,** 996-1006.

Kenyon, S., Boulvain, M. & Neilson, J. P. 2013. Antibiotics for preterm rupture of membranes. *Cochrane Database Syst Rev,* 12**,** CD001058.

Khalil, A., Rezende, J., Akolekar, R., Syngelaki, A. & Nicolaides, K. H. 2013. Maternal racial origin and adverse pregnancy outcome: a cohort study. *Ultrasound Obstet Gynecol,* 41**,** 278-85.

Kim, C. J., Romero, R., Chaemsaithong, P., Chaiyasit, N., Yoon, B. H. & Kim, Y. M. 2015. Acute chorioamnionitis and funisitis: definition, pathologic features, and clinical significance. *Am J Obstet Gynecol,* 213**,** S29-52.

Kim, M. C., Kim, C. M., Kang, O. C., Zhang, Y., Liu, Z., Wangmu, D., Wei, Z., Huang, Y. & Peng, F. 2016. Hymenobacter rutilus sp. nov., isolated from marine sediment in Arctic. *Int J Syst Evol Microbiol*.21.

Kindinger, L. M., Macintyre, D. A., Lee, Y. S., Marchesi, J. R., Smith, A., Mcdonald, J. A., Terzidou, V., Cook, J. R., Lees, C., Israfil-Bayli, F., Faiza, Y., Toozs-Hobson, P., Slack, M., Cacciatore, S., Holmes, E., Nicholson, J. K., Teoh, T. G. & Bennett, P. R. 2016. Relationship between vaginal microbial dysbiosis, inflammation, and pregnancy outcomes in cervical cerclage. *Sci Transl Med,* 8**,** 350ra102.

Kircher, M., Sawyer, S. & Meyer, M. 2011. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research,* 40**,** e3-e3.

Klebanoff, M. & Searle, K. 2006. The role of inflammation in preterm birth—focus on periodontitis. *BJOG: An International Journal of Obstetrics & Gynaecology,* 113**,** 43-45.

Kliman, H. J. 2014. Comment on "The placenta harbors a unique microbiome". *Science Translational Medicine,* 6**,** 254le4.

Ko, K. S., Lee, N. Y., Oh, W. S., Lee, J. H., Ki, H. K., Peck, K. R. & Song, J. H. 2005. Tepidimonas arfidensis Sp. Nov., a Novel Gram-negative and thermophilic bacterium isolated from the bone marrow of a patient with leukemia in Korea. *Microbiol Immunol,* 49**,** 785-8.

Kock, K., Kock, F., Klein, K., Bancher-Todesca, D. & Helmer, H. 2010. Diabetes mellitus and the risk of preterm birth with regard to the risk of spontaneous preterm birth. *J Matern Fetal Neonatal Med,* 23**,** 1004-8.

Kozuki, N., Lee, A. C., Silveira, M. F., Sania, A., Vogel, J. P., Adair, L., Barros, F., Caulfield, L. E., Christian, P., Fawzi, W., Humphrey, J., Huybregts, L., Mongkolchati, A., Ntozini, R., Osrin, D., Roberfroid, D., Tielsch, J., Vaidya, A., Black, R. E. & Katz, J. 2013. The associations of parity and maternal age with small-for-gestational-age, preterm, and neonatal and infant mortality: a meta-analysis. *BMC Public Health,* 13**,** 1-10.

Kyrklund-Blomberg, N. B. & Cnattingius, S. 1998. Preterm birth and maternal smoking: risks related to gestational age and onset of delivery. *Am J Obstet Gynecol,* 179**,** 1051-5.

Lauder, A. P., Roche, A. M., Sherrill-Mix, S., Bailey, A., Laughlin, A. L., Bittinger, K., Leite, R., Elovitz, M. A., Parry, S. & Bushman, F. D. 2016. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome,* 4**,** 29.

Lawn, J. E., Kerber, K., Enweronu-Laryea, C. & Cousens, S. 2010. 3.6 million neonatal deaths--what is progressing and what is not? *Semin Perinatol,* 34**,** 371-86.

Lee, J. J., Park, S. J., Lee, Y. H., Lee, S. Y., Ten, L. N. & Jung, H. Y. 2017. Hymenobacter aquaticus sp. nov., a radiation-resistant bacterium isolated from the Han River. *Int J Syst Evol Microbiol*.

Leighton, J. D. & Pfeilschifter, J. 1990. Interleukin 1- and tumor necrosis factor-stimulation of prostaglandin E2 synthesis in MDCK cells, and potentiation of this effect by cycloheximide. *FEBS Letters,* 259**,** 289-292.

Leitich, H., Bodner-Adler, B., Brunbauer, M., Kaider, A., Egarter, C. & Husslein, P. 2003. Bacterial vaginosis as a risk factor for preterm delivery: A meta-analysis. *American Journal of Obstetrics and Gynecology,* 189**,** 139-147.

Leon, L. J., Solanky, N., Stalman, S. E., Demetriou, C., Abu-Amero, S., Stanier, P., Regan, L. & Moore, G. E. 2016. A new biological and clinical resource for research into pregnancy complications: The Baby Bio Bank. *Placenta,* 46**,** 31-37.

Liggins, G. C. 1981. Cervical ripening as an inflammatory reaction. . *In:* ELLWOOD, D. A. A. A., A.B.M. (ed.) *he Cervix in Pregnancy and Labour: Clinical and Biochemical Investigations.* Edinburgh: Churchill Livingstone.

Liu, L., Oza, S., Hogan, D., Perin, J., Rudan, I., Lawn, J. E., Cousens, S., Mathers, C. & Black, R. E. 2015. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. *Lancet,* 385**,** 430-40.

Lockwood, C. J., Senyei, A. E., Dische, M. R., Casal, D., Shah, K. D., Thung, S. N., Jones, L., Deligdisch, L. & Garite, T. J. 1991. Fetal fibronectin in cervical and vaginal secretions as a predictor of preterm delivery. *N Engl J Med,* 325**,** 669-74.

Lopez, E., Raymond, J., Patkai, J., Ayoubl, M. E., Schmitz, T., Moriette, G. & Jarreau, P. H. 2010a. Capnocytophaga species and preterm birth: case series and review of the literature. *Clinical Microbiology and Infection,* 16**,** 1539-1543.

Lopez, E., Raymond, J., Patkai, J., El Ayoubi, M., Schmitz, T., Moriette, G. & Jarreau, P. H. 2010b. Capnocytophaga species and preterm birth: case series and review of the literature. *Clin Microbiol Infect,* 16**,** 1539-43.

Love, M. I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

Lozupone, C. & Knight, R. 2005. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology,* 71**,** 8228-8235.

Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol,* 73**,** 1576-85.

Lyon, C. J., Law, R. E. & Hsueh, W. A. 2003. Minireview: adiposity, inflammation, and atherogenesis. *Endocrinology,* 144**,** 2195-200.

Macintyre, D. A., Chandiramani, M., Lee, Y. S., Kindinger, L., Smith, A., Angelopoulos, N., Lehne, B., Arulkumaran, S., Brown, R., Teoh, T. G., Holmes, E., Nicoholson, J. K., Marchesi, J. R. & Bennett, P. R. 2015. The vaginal microbiome during pregnancy and the postpartum period in a European population. *Scientific Reports,* 5**,** 8988.

Mahé, F., Rognes, T., Quince, C., De Vargas, C. & Dunthorn, M. 2014. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ,* 2**,** e593.

Mangham, L. J., Petrou, S., Doyle, L. W., Draper, E. S. & Marlow, N. 2009. The cost of preterm birth throughout childhood in England and Wales. *Pediatrics,* 123**,** e312-27.

Marconi, C., De Andrade Ramos, B. R., Peraçoli, J. C., Donders, G. G. G. & Da Silva, M. G. 2011. Amniotic Fluid Interleukin-1 Beta and Interleukin-6, but not Interleukin-8 Correlate with Microbial Invasion of the Amniotic Cavity in Preterm Labor. *American Journal of Reproductive Immunology,* 65**,** 549-556.

Markenson, G. R., Martin, R. K., Tillotson-Criss, M., Foley, K. S., Stewart, R. S., Jr. & Yancey, M. 1997. The use of the polymerase chain reaction to detect bacteria in amniotic fluid in pregnancies complicated by preterm labor. *Am J Obstet Gynecol,* 177**,** 1471-7.

Martin, J. A., Hamilton, B. E. & Osterman, M. J. 2014. Births in the United States, 2013. *NCHS Data Brief***,** 1-8.

Martin, J. A. & Osterman, M. J. 2013. Preterm births - United States, 2006 and 2010. *MMWR Surveill Summ,* 62 Suppl 3**,** 136-8.

Matsuura, H., Takio, K., Titani, K., Greene, T., Levery, S. B., Salyan, M. E. & Hakomori, S. 1988. The oncofetal structure of human fibronectin defined by monoclonal antibody FDC-6. Unique structural requirement for the antigenic specificity provided by a glycosylhexapeptide. *J Biol Chem,* 263**,** 3314-22.

Mcdonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., Desantis, T. Z. & Probst, A. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J,* 6.

Mcmurdie, P. J. & Holmes, S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE,* 8**,** e61217.

Mcmurdie, P. J. & Holmes, S. 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput Biol,* 10**,** e1003531.

Meis, P. J. & Aleman, A. 2004. Progesterone treatment to prevent preterm birth. *Drugs,* 64**,** 2463-74.

Meis, P. J., Klebanoff, M., Thom, E., Dombrowski, M. P., Sibai, B., Moawad, A. H., Spong, C. Y., Hauth, J. C., Miodovnik, M., Varner, M. W., Leveno, K. J., Caritis, S. N., Iams, J. D., Wapner, R. J., Conway, D., O'sullivan, M. J., Carpenter, M., Mercer, B., Ramin, S. M., Thorp, J. M., Peaceman, A. M. & Gabbe, S. 2003. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *N Engl J Med,* 348**,** 2379-85.

Meis, P. J., Michielutte, R., Peters, T. J., Wells, H. B., Sands, R. E., Coles, E. C. & Johns, K. A. 1995. Factors associated with preterm birth in Cardiff, Wales. I. Univariable and multivariable analysis. *Am J Obstet Gynecol,* 173**,** 590-6.

Menon, R., Peltier, M. R., Eckardt, J. & Fortunato, S. J. 2009. Diversity in cytokine response to bacteria associated with preterm birth by fetal membranes. *American Journal of Obstetrics and Gynecology,* 201**,** 306.e1-306.e6.

Menon, R., Torloni, M. R., Voltolini, C., Torricelli, M., Merialdi, M., Betran, A. P., Widmer, M., Allen, T., Davydova, I., Khodjaeva, Z., Thorsen, P., Kacerovsky, M., Tambor, V., Massinen, T., Nace, J. & Arora, C. 2011. Biomarkers of spontaneous preterm birth: an overview of the literature in the last four decades. *Reprod Sci,* 18**,** 1046-70.

Mercer, B. M., Goldenberg, R. L., Moawad, A. H., Meis, P. J., Iams, J. D., Das, A. F., Caritis, S. N., Miodovnik, M., Menard, M. K., Thurnau, G. R., Dombrowski, M. P., Roberts, J. M. & Mcnellis, D. 1999. The preterm prediction study: effect of gestational age and cause of preterm birth on subsequent obstetric outcome. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. *Am J Obstet Gynecol,* 181**,** 1216-21.

Mestan, K., Yu, Y., Thorsen, P., Skogstrand, K., Matoba, N., Liu, X., Kumar, R., Hougaard, D. M., Gupta, M., Pearson, C., Ortiz, K., Bauchner, H. & Wang, X. 2009. Cord blood biomarkers of the fetal inflammatory response. *J Matern Fetal Neonatal Med,* 22**,** 379-87.

Migale, R., Herbert, B. R., Lee, Y. S., Sykes, L., Waddington, S. N., Peebles, D., Hagberg, H., Johnson, M. R., Bennett, P. R. & Macintyre, D. A. 2015. Specific Lipopolysaccharide Serotypes Induce Differential Maternal and Neonatal Inflammatory Responses in a Murine Model of Preterm Labor. *Am J Pathol,* 185**,** 2390-401.

Migale, R., Macintyre, D. A., Cacciatore, S., Lee, Y. S., Hagberg, H., Herbert, B. R., Johnson, M. R., Peebles, D., Waddington, S. N. & Bennett, P. R. 2016. Modeling hormonal and inflammatory contributions to preterm and term labor using uterine temporal transcriptomics. *BMC Med,* 14**,** 86.

Mitchell, M. D., Trautman, M. S. & Dudley, D. J. 1993. Cytokine networking in the placenta. *Placenta,* 14**,** 249-75.

Mohan, A. R., Loudon, J. A. & Bennett, P. R. 2004. Molecular and biochemical mechanisms of preterm labour. *Seminars in Fetal and Neonatal Medicine,* 9**,** 437-444.

Moles, L., Gómez, M., Heilig, H., Bustos, G., Fuentes, S., De Vos, W., Fernández, L., Rodríguez, J. M. & Jiménez, E. 2013. Bacterial Diversity in Meconium of Preterm Neonates and Evolution of Their Fecal Microbiota during the First Month of Life. *PLOS ONE,* 8**,** e66986.

Molnár, M., Romero, R. & Hertelendy, F. 1993. Interleukin-1 and tumor necrosis factor stimulate arachidonic acid release and phospholipid metabolism in human myometrial cells. *American Journal of Obstetrics and Gynecology,* 169**,** 825-829.

Moore, G. E., Ishida, M., Demetriou, C., Al-Olabi, L., Leon, L. J., Thomas, A. C., Abu-Amero, S., Frost, J. M., Stafford, J. L., Chaoqun, Y., Duncan, A. J., Baigel, R., Brimioulle, M., Iglesias-Platas, I., Apostolidou, S., Aggarwal, R., Whittaker, J. C., Syngelaki, A., Nicolaides, K. H., Regan, L., Monk, D. & Stanier, P. 2015. The role and interaction of imprinted genes in human fetal growth. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences,* 370**,** 20140074.

Morales, S. E. & Holben, W. E. 2009. Empirical Testing of 16S rRNA Gene PCR Primer Pairs Reveals Variance in Target Specificity and Efficacy Not Suggested by In Silico Analysis. *Applied and Environmental Microbiology,* 75**,** 2677-2683.

Moreira, C., Rainey, F. A., Nobre, M. F., Da Silva, M. T. & Da Costa, M. S. 2000. Tepidimonas ignava gen. nov., sp. nov., a new chemolithoheterotrophic and slightly thermophilic member of the beta-Proteobacteria. *Int J Syst Evol Microbiol,* 50 Pt 2**,** 735-42.

Moser, K., Stanfield, K. M. & Leon, D. A. 2008. Birthweight and gestational age by ethnic group, England and Wales 2005: introducing new data on births. *Health Stat Q***,** 22-31, 34-55.

Mueller-Heubach, E., Rubinstein, D. N. & Schwarz, S. S. 1990. Histologic chorioamnionitis and preterm delivery in different patient populations. *Obstet Gynecol,* 75**,** 622-6.

Muglia, L. J. & Katz, M. 2010. The Enigma of Spontaneous Preterm Birth. *New England Journal of Medicine,* 362**,** 529-535.

Mund, M., Louwen, F., Klingelhoefer, D. & Gerber, A. 2013. Smoking and Pregnancy — A Review on the First Major Environmental Risk Factor of the Unborn. *International Journal of Environmental Research and Public Health,* 10**,** 6485-6499.

Musilova, I., Pliskova, L., Kutova, R., Hornychova, H., Jacobsson, B. & Kacerovsky, M. 2014. Ureaplasma species and Mycoplasma hominis in cervical fluid of pregnancies complicated by preterm prelabor rupture of membranes. *J Matern Fetal Neonatal Med***,** 1-7.

Mussalli, G. M., Blanchard, R., Brunnert, S. R. & Hirsch, E. 1999. Inflammatory Cytokines in a Murine Model of Infection-Induced Preterm Labor: Cause or Effect? *Journal of the Society for Gynecologic Investigation,* 6**,** 188-195.

Mysorekar, I. U. & Cao, B. 2014. Microbiome in Parturition and Preterm Birth. *Semin Reprod Med,* 32**,** 050-055.

Nakazawa, F., Sato, M., Poco, S. E., Hashimura, T., Ikeda, T., Kalfas, S., Sundqvist, G. & Hoshino, E. 2000. Description of Mogibacterium pumilum gen. nov.,

sp. nov. and Mogibacterium vescum gen. nov., sp. nov., and reclassification of Eubacterium timidum (Holdeman et al. 1980) as Mogibacterium timidum gen. nov., comb. nov. *Int J Syst Evol Microbiol,* 50 Pt 2**,** 679-88.

Nelissen, E. C. M., Van Montfoort, A. P. A., Dumoulin, J. C. M. & Evers, J. L. H. 2011. Epigenetics and the placenta. *Human Reproduction Update,* 17**,** 397-417.

Offenbacher, S., Boggess, K. A., Murtha, A. P., Jared, H. L., Lieff, S., Mckaig, R. G., Mauriello, S. M., Moss, K. L. & Beck, J. D. 2006. Progressive periodontal disease and risk of very preterm delivery. *Obstet Gynecol,* 107**,** 29-36.

Offenbacher, S., Katz, V., Fertik, G., Collins, J., Boyd, D., Maynor, G., Mckaig, R. & Beck, J. 1996. Periodontal Infection as a Possible Risk Factor for Preterm Low Birth Weight. *Journal of Periodontology,* 67**,** 1103-1113.

Offenbacher, S., Lieff, S., Boggess, K. A., Murtha, A. P., Madianos, P. N., Champagne, C. M., Mckaig, R. G., Jared, H. L., Mauriello, S. M., Auten, R. L., Jr., Herbert, W. N. & Beck, J. D. 2001. Maternal periodontitis and prematurity. Part I: Obstetric outcome of prematurity and growth restriction. *Ann Periodontol,* 6**,** 164-74.

Office for National Statistics 2014. Gestation-specific infant mortality, 2012.

Office for National Statistics 2015. Pregnancy and ethnic factors influencing births and infant mortality: 2013.

Oh, K. J., Lee, K. A., Sohn, Y. K., Park, C. W., Hong, J. S., Romero, R. & Yoon, B. H. 2010. Intraamniotic infection with genital mycoplasmas exhibits a more intense inflammatory response than intraamniotic infection with other microorganisms in patients with preterm premature rupture of membranes. *Am J Obstet Gynecol,* 203**,** 211.e1-8.

Ollberding, N. J., Völgyi, E., Macaluso, M., Kumar, R., Morrow, C., Tylavsky, F. A. & Piyathilake, C. J. 2016. Urinary Microbiota Associated with Preterm Birth: Results from the Conditions Affecting Neurocognitive Development and Learning in Early Childhood (CANDLE) Study. *PLOS ONE,* 11**,** e0162302.

Olson, D. M. 2003. The role of prostaglandins in the initiation of parturition. *Best Practice and Research: Clinical Obstetrics and Gynaecology,* 17**,** 717-730.

Onderdonk, A. B., Delaney, M. L., Dubois, A. M., Allred, E. N., Leviton, A. & Extremely Low Gestational Age Newborns Study, I. 2008a. Detection of bacteria in placental tissues obtained from extremely low gestational age neonates. *Am J Obstet Gynecol,* 198**,** 110 e1-7.

Onderdonk, A. B., Hecht, J. L., Mcelrath, T. F., Delaney, M. L., Allred, E. N. & Leviton, A. 2008b. Colonization of second-trimester placenta parenchyma. *American Journal of Obstetrics and Gynecology,* 199**,** 52.e1-52.e10.

Payne, M. S. & Bayatibojakhi, S. 2014. Exploring preterm birth as a polymicrobial disease: an overview of the uterine microbiome. *Front Immunol,* 5**,** 595.

Pelzer, E., Gomez-Arango, L. F., Barrett, H. L. & Nitert, M. D. 2016. Maternal health and the placental microbiome. *Placenta*.

Petraglia, F., Imperatore, A. & Challis, J. R. 2010. Neuroendocrine mechanisms in pregnancy and parturition. *Endocr Rev,* 31**,** 783-816.

Plotkin, S. A. 2006. The History of Rubella and Rubella Vaccination Leading to Elimination. *Clinical Infectious Diseases,* 43**,** S164-S168.

Plunkett, J. & Muglia, L. J. 2008. Genetic contributions to preterm birth: implications from epidemiological and genetic association studies. *Ann Med,* 40**,** 167-95.

Pollack, R. N. & Divon, M. Y. 1992. Intrauterine growth retardation: definition, classification, and etiology. *Clinical Obstetetrics and Gynecology,* 35**,** 99-107.

Price, M. N., Dehal, P. S. & Arkin, A. P. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE,* 5**,** e9490.

Prince, A. L., Ma, J., Kannan, P. S., Alvarez, M., Gisslen, T., Harris, R. A., Sweeney, E. L., Knox, C. L., Lambers, D. S., Jobe, A. H., Chougnet, C. A., Kallapur, S. G. & Aagaard, K. M. 2016. The placental membrane microbiome is altered among subjects with spontaneous preterm birth with and without chorioamnionitis. *Am J Obstet Gynecol,* 214**,** 627.e1-627.e16.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. & Glöckner, F. O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res,* 35**,** 7188-96.

Rai, R. & Regan, L. 2006. Recurrent miscarriage. *The Lancet,* 368**,** 601-11.

Rantakokko-Jalava, K. & Jalava, J. 2002. Optimal DNA isolation method for detection of bacteria in clinical specimens by broad-range PCR. *J Clin Microbiol,* 40**,** 4211-7.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S. & Mcculle, S. L. 2011. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci USA,* 108.

Rcog 2011. Cervical Cerclage. *Green-top Guideline No.60.* Royal College of Obstetricians & Gynaecologists.

Reddick, K. L., Jhaveri, R., Gandhi, M., James, A. H. & Swamy, G. K. 2011. Pregnancy outcomes associated with viral hepatitis. *J Viral Hepat,* 18**,** e394-8.

Redline, R. W. 2004. Placental inflammation. *Semin Neonatol,* 9**,** 265-74.

Regan, L., Braude, P. R. & Trembath, P. L. 1989. Influence of past reproductive performance on risk of spontaneous abortion. *BMJ,* 299**,** 541-5.

Restrepo-Mendez, M. C., Lawlor, D. A., Horta, B. L., Matijasevich, A., Santos, I. S., Menezes, A. M., Barros, F. C. & Victora, C. G. 2015. The association of maternal age with birthweight and gestational age: a cross-cohort comparison. *Paediatr Perinat Epidemiol,* 29**,** 31-40.

Rinke, C., Low, S., Woodcroft, B. J., Raina, J.-B., Skarshewski, A., Le, X. H., Butler, M. K., Stocker, R., Seymour, J., Tyson, G. W. & Hugenholtz, P. 2016. Validation of picogram- and femtogram-input DNA libraries for microscale metagenomics. *PeerJ,* 4**,** e2486.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res,* 43**,** e47.

Roberts, C. L., Ford, J. B., Algert, C. S., Antonsen, S., Chalmers, J., Cnattingius, S., Gokhale, M., Kotelchuck, M., Melve, K. K., Langridge, A., Morris, C., Morris, J. M., Nassar, N., Norman, J. E., Norrie, J., Sørensen, H. T., Walker, R. & Weir, C. J. 2011. Population-based trends in pregnancy hypertension and pre-eclampsia: an international comparative study. *BMJ Open,* 1.

Robertson, S. A., Skinner, R. J. & Care, A. S. 2006. Essential role for IL-10 in resistance to lipopolysaccharide-induced preterm labor in mice. *J Immunol,* 177**,** 4888-96.

Romano-Keeler, J. & Weitkamp, J.-H. 2014. Maternal influences on fetal microbial colonization and immune development. *Pediatr Res*.

Romero, R., Dey, S. K. & Fisher, S. J. 2014a. Preterm labor: one syndrome, many causes. *Science,* 345**,** 760-5.

Romero, R., Espinoza, J., Kusanovic, J. P., Gotsch, F., Hassan, S., Erez, O., Chaiworapongsa, T. & Mazor, M. 2006. The preterm parturition syndrome. *BJOG,* 113 Suppl 3**,** 17-42.

Romero, R., Hassan, S. S., Gajer, P., Tarca, A. L., Fadrosh, D. W., Bieda, J., Chaemsaithong, P., Miranda, J., Chaiworapongsa, T. & Ravel, J. 2014b. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome,* 2**,** 18.

Romero, R., Hassan, S. S., Gajer, P., Tarca, A. L., Fadrosh, D. W., Nikita, L., Galuppi, M., Lamont, R. F., Chaemsaithong, P., Miranda, J., Chaiworapongsa, T. & Ravel, J. 2014c. The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome,* 2**,** 1-19.

Romero, R., Mazor, M., Brandt, F., Sepulveda, W., Avila, C., Cotton, D. B. & Dinarello, C. A. 1992. Interleukin-1 alpha and interleukin-1 beta in preterm and term human parturition. *Am J Reprod Immunol,* 27**,** 117-23.

Romero, R., Mazor, M. & Tartakovsky, B. 1991. Systemic administration of interleukin-1 induces preterm parturition in mice. *Am J Obstet Gynecol,* 165**,** 969-71.

Romero, R., Miranda, J., Chaiworapongsa, T., Chaemsaithong, P., Gotsch, F., Dong, Z., Ahmed, A. I., Yoon, B. H., Hassan, S. S., Kim, C. J., Korzeniewski, S. J. & Yeo, L. 2014d. A novel molecular microbiologic technique for the rapid diagnosis of microbial invasion of the amniotic cavity and intra-amniotic infection in preterm labor with intact membranes. *Am J Reprod Immunol,* 71**,** 330-58.

Romero, R. & Tartakovsky, B. 1992. The natural interleukin-1 receptor antagonist prevents interleukin-1-induced preterm delivery in mice. *Am J Obstet Gynecol,* 167**,** 1041-5.

Rosselli, R., Romoli, O., Vitulo, N., Vezzi, A., Campanaro, S., De Pascale, F., Schiavon, R., Tiarca, M., Poletto, F., Concheri, G., Valle, G. & Squartini, A. 2016. Direct 16S rRNA-seq from bacterial communities: a PCR-independent approach to simultaneously assess microbial diversity and functional activity potential of each taxon. *Scientific Reports,* 6**,** 32165.

Rours, G. I., De Krijger, R. R., Ott, A., Willemse, H. F., De Groot, R., Zimmermann, L. J., Kornelisse, R. F., Verbrugh, H. A. & Verkooijen, R. P. 2011. Chlamydia trachomatis and placental inflammation in early preterm delivery. *Eur J Epidemiol,* 26**,** 421-8.

Royston, J. P. 1982. An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 31**,** 115-124.

Rstudioteam 2015. RStudio: Integrated Development for R. 0.99.484 ed.: RStudio, Inc., Boston, MA

Saigal, S. & Doyle, L. W. 2008. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet,* 371**,** 261-269.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O. & Moffatt, M. F. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol,* 12.

Satokari, R., Gronroos, T., Laitinen, K., Salminen, S. & Isolauri, E. 2009. Bifidobacterium and Lactobacillus DNA in the human placenta. *Lett Appl Microbiol,* 48**,** 8-12.

Schaaf, J. M., Liem, S. M., Mol, B. W., Abu-Hanna, A. & Ravelli, A. C. 2013. Ethnic and racial disparities in the risk of preterm birth: a systematic review and meta-analysis. *Am J Perinatol,* 30**,** 433-50.

Schlafer, D. H., Yuh, B., Foley, G. L., Elssaser, T. H., Sadowsky, D. & Nathanielsz, P. W. 1994. Effect of Salmonella endotoxin administered to the pregnant sheep at 133-142 days gestation on fetal oxygenation, maternal and fetal adrenocorticotropic hormone and cortisol, and maternal plasma tumor necrosis factor alpha concentrations. *Biology of Reproduction,* 50**,** 1297-1302.

Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. 2012. PCR inhibitors – occurrence, properties and removal. *Journal of Applied Microbiology,* 113**,** 1014-1026.

Seematter, G., Binnert, C., Martin, J. L. & Tappy, L. 2004. Relationship between stress, inflammation and metabolism. *Curr Opin Clin Nutr Metab Care,* 7**,** 169-73.

Segal, L. N., Alekseyenko, A. V., Clemente, J. C., Kulkarni, R., Wu, B. & Chen, H. 2013. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome.,* 1.

Sergeant, M. J., Constantinidou, C., Cogan, T., Penn, C. W. & Pallen, M. J. 2012. High-Throughput Sequencing of 16S rRNA Gene Amplicons: Effects of Extraction Procedure, Primer Length and Annealing Temperature. *PLoS One,* 7.

Seri, I. & Evans, J. 2008. Limits of viability: definition of the gray zone. *Journal of Perinatology,* 28**,** S4-S8.

Shannon, C. E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal,* 27**,** 623-656.

Shaw, G. M., Wise, P. H., Mayo, J., Carmichael, S. L., Ley, C., Lyell, D. J., Shachar, B. Z., Melsop, K., Phibbs, C. S., Stevenson, D. K., Parsonnet, J. & Gould, J. B. 2014. Maternal prepregnancy body mass index and risk of spontaneous preterm birth. *Paediatr Perinat Epidemiol,* 28**,** 302-11.

Shobokshi, A. & Shaarawy, M. 2002. Maternal serum and amniotic fluid cytokines in patients with preterm premature rupture of membranes with and without intrauterine infection. *Int J Gynaecol Obstet,* 79**,** 209-15.

Silva, P. E., Costa, P. S., Ávila, M. P., Suhadolnik, M. L. S., Reis, M. P., Salgado, A. P. C., Lima, M. F., Chartone-Souza, E. & Nascimento, A. M. 2015. Leprous lesion presents enrichment of opportunistic pathogenic bacteria. *SpringerPlus,* 4**,** 187.

Simhan, H. N., Krohn, M. A., Roberts, J. M., Zeevi, A. & Caritis, S. N. 2003. Interleukin-6 promoter -174 polymorphism and spontaneous preterm birth. *Am J Obstet Gynecol,* 189**,** 915-8.

Simmons, R. A. 2009. Developmental origins of adult disease. *Pediatr Clin North Am,* 56**,** 449-66.

Sisti, G., Kanninen, T. T. & Witkin, S. S. 2016. Maternal immunity and pregnancy outcome: focus on preconception and autophagy. *Genes Immun,* 17**,** 1-7.

Slyker, J. A., Patterson, J., Ambler, G., Richardson, B. A., Maleche-Obimbo, E., Bosire, R., Mbori-Ngacha, D., Farquhar, C. & John-Stewart, G. 2014. Correlates and outcomes of preterm birth, low birth weight, and small for gestational age in HIV-exposed uninfected infants. *BMC Pregnancy Childbirth,* 14**,** 7.

Srikhajon, K., Shynlova, O., Preechapornprasert, A., Chanrachakul, B. & Lye, S. 2014. A new role for monocytes in modulating myometrial inflammation during human labor. *Biol Reprod,* 91**,** 10.

Staley, J. T. & Konopka, A. 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol,* 39**,** 321-46.

Steel, J. H., Malatos, S., Kennea, N., Edwards, A. D., Miles, L., Duggan, P., Reynolds, P. R., Feldman, R. G. & Sullivan, M. H. 2005. Bacteria and inflammatory cells in fetal membranes do not always cause preterm labor. *Pediatr Res,* 57**,** 404-11.

Steinhoff, M. C., Omer, S. B., Roy, E., El Arifeen, S., Raqib, R., Dodd, C., Breiman, R. F. & Zaman, K. 2012. Neonatal outcomes after influenza immunization during pregnancy: a randomized controlled trial. *Cmaj,* 184**,** 645-53.

Stirrat, G. M. 1990. Recurrent miscarriage. *The Lancet,* 336**,** 673-5.

Stoll, B. J., Hansen, N. I., Bell, E. F., Walsh, M. C., Carlo, W. A., Shankaran, S., Laptook, A. R., Sanchez, P. J., Van Meurs, K. P., Wyckoff, M., Das, A., Hale, E. C., Ball, M. B., Newman, N. S., Schibler, K., Poindexter, B. B., Kennedy, K. A., Cotten, C. M., Watterberg, K. L., D'angio, C. T., Demauro, S. B., Truog, W. E., Devaskar, U. & Higgins, R. D. 2015. Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993-2012. *Jama,* 314**,** 1039-51.

Stout, M. J., Conlon, B., Landeau, M., Lee, I., Bower, C., Zhao, Q., Roehl, K. A., Nelson, D. M., Macones, G. A. & Mysorekar, I. U. 2013. Identification of intracellular bacteria in the basal plate of the human placenta in term and preterm gestations. *American Journal of Obstetrics and Gynecology,* 208**,** 226.e1-226.e7.

Subramaniam, A., Abramovici, A., Andrews, W. W. & Tita, A. T. 2012. Antimicrobials for Preterm Birth Prevention: An Overview. *Infectious Diseases in Obstetrics and Gynecology,* 2012**,** 12.

Sykes, L., Macintyre, D. A., Yap, X. J., Teoh, T. G. & Bennett, P. R. 2012. The Th1:th2 dichotomy of pregnancy and preterm labour. *Mediators Inflamm,* 2012**,** 967629.

Tency, I. 2014. Inflammatory response in maternal serum during preterm labour. *Facts, Views & Vision in ObGyn,* 6**,** 19-30.

Torbe, A., Czajka, R., Kordek, A., Rzepka, R., Kwiatkowski, S. & Rudnicki, J. 2007. Maternal serum proinflammatory cytokines in preterm labor with intact membranes: neonatal outcome and histological associations. *Eur Cytokine Netw,* 18**,** 102-7.

Torloni, M. R., Betran, A. P., Daher, S., Widmer, M., Dolan, S. M., Menon, R., Bergel, E., Allen, T. & Merialdi, M. 2009. Maternal BMI and preterm birth: a

systematic review of the literature with meta-analysis. *J Matern Fetal Neonatal Med,* 22**,** 957-70.

Tornblom, S. A., Klimaviciute, A., Bystrom, B., Chromek, M., Brauner, A. & Ekman-Ordeberg, G. 2005. Non-infected preterm parturition is related to increased concentrations of IL-6, IL-8 and MCP-1 in human cervix. *Reprod Biol Endocrinol,* 3**,** 39.

Tribe, R. M., Moriarty, P., Dalrymple, A., Hassoni, A. A. & Poston, L. 2003. Interleukin-1beta induces calcium transients and enhances basal and store operated calcium entry in human myometrial smooth muscle. *Biol Reprod,* 68**,** 1842-9.

Tringe, S. G. & Hugenholtz, P. 2008. A renaissance for the pioneering 16S rRNA gene. *Current Opinion in Microbiology,* 11**,** 442-446.

Uneke, C. J. 2007. Impact of placental Plasmodium falciparum malaria on pregnancy and perinatal outcome in sub-Saharan Africa: I: introduction to placental malaria. *Yale J Biol Med,* 80**,** 39-50.

Verstraelen, H., Vilchez-Vargas, R., Desimpel, F., Jauregui, R., Vankeirsbilck, N., Weyers, S., Verhelst, R., De Sutter, P., Pieper, D. H. & Van De Wiele, T. 2016. Characterisation of the human uterine microbiome in non-pregnant women through deep sequencing of the V1-2 region of the 16S rRNA gene. *PeerJ,* 4**,** e1602.

Vitoratos, N., Mastorakos, G., Kountouris, A., Papadias, K. & Creatsas, G. 2007. Positive association of serum interleukin-1beta and CRH levels in women with pre-term labor. *J Endocrinol Invest,* 30**,** 35-40.

Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology,* 73**,** 5261-5267.

Wang, X., Buhimschi, C. S., Temoin, S., Bhandari, V., Han, Y. W. & Buhimschi, I. A. 2013. Comparative microbial analysis of paired amniotic fluid and cord blood from pregnancies complicated by preterm birth and early-onset neonatal sepsis. *PLoS One,* 8**,** e56131.

Wang, Y. & Qian, P.-Y. 2009. Conservative Fragments in Bacterial 16S rRNA Genes and Primer Design for 16S Ribosomal DNA Amplicons in Metagenomic Studies. *PLoS ONE,* 4**,** e7401.

Wassenaar, T. M. & Panigrahi, P. 2014. Is a foetus developing in a sterile environment? *Lett Appl Microbiol,* 59**,** 572-9.

Watari, M., Watari, H., Disanto, M. E., Chacko, S., Shi, G.-P. & Strauss, J. F. 1999. Pro-Inflammatory Cytokines Induce Expression of Matrix-Metabolizing Enzymes in Human Cervical Smooth Muscle Cells. *The American Journal of Pathology,* 154**,** 1755-1762.

Weiss, S., Amir, A., Hyde, E. R., Metcalf, J. L., Song, S. J. & Knight, R. 2014. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol.,* 15.

Wesolowska-Andersen, A., Bahl, M. I., Carvalho, V., Kristiansen, K., Sicheritz-Pontén, T., Gupta, R. & Licht, T. R. 2014. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome,* 2**,** 1-11.

Who 1988. Geographic variation in the incidence of hypertension in pregnancy. World Health Organization International Collaborative Study of

Hypertensive Disorders of Pregnancy. *American Journal of Obstetrics and Gynecology,* 158**,** 80-3.

Who 2011. WHO recommendations for prevention and treatment of pre-eclampsia and eclampsia. Geneva: World Health Organization.

Wickham, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*, Springer New York.

Wilcox, A. J., Skjaerven, R. & Lie, R. T. 2008. Familial patterns of preterm delivery: maternal and fetal contributions. *Am J Epidemiol,* 167**,** 474-9.

Wilkins-Haug, L., Quade, B. & Morton, C. C. 2006. Confined placental mosaicism as a risk factor among newborns with fetal growth restriction. *Prenatal Diagnosis,* 26**,** 428-32.

Williams, A. E. 2012. *Immunology: Mucosal and Body Surface Defences*, Wiley-Blackwell.

Willner, D., Daly, J., Whiley, D., Grimwood, K., Wainwright, C. E. & Hugenholtz, P. 2012. Comparison of DNA Extraction Methods for Microbial Community Profiling with an Application to Pediatric Bronchoalveolar Lavage Samples. *PLoS One,* 7.

Witkin, S. S. 2014. The vaginal microbiome, vaginal anti-microbial defence mechanisms and the clinical challenge of reducing infection-related preterm birth. *BJOG: An International Journal of Obstetrics & Gynaecology***,** 122(2):213-8

Witkin, S. S., Linhares, I. M. & Giraldo, P. 2007. Bacterial flora of the female genital tract: function and immune regulation. *Best Pract Res Clin Obstet Gynaecol,* 21.

Woese, C. R. & Fox, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A,* 74**,** 5088-90.

Wolak, M. E., Fairbairn, D. J. & Paulsen, Y. R. 2012. Guidelines for estimating repeatability. *Methods in Ecology and Evolution,* 3**,** 129-137.

Wu, G. D. & Lewis, J. D. 2013. Analysis of the human gut microbiome and association with disease. *Clin Gastroenterol Hepatol,* 11**,** 774-7.

Wu, J., Peters, B. A., Dominianni, C., Zhang, Y., Pei, Z., Yang, L., Ma, Y., Purdue, M. P., Jacobs, E. J., Gapstur, S. M., Li, H., Alekseyenko, A. V., Hayes, R. B. & Ahn, J. 2016. Cigarette smoking and the oral microbiome in a large study of American adults. *ISME J*.

Yang, Q., El Sayed, Y., Shaw, G. M., Fu, J., Schilling, J. & Madan, A. 2010. Second-trimester serum cytokines in women who develop spontaneous preterm labor at less than 28 weeks' gestation versus term labor. *Am J Perinatol,* 27**,** 31-6.

Young, A., Thomson, A. J., Ledingham, M., Jordan, F., Greer, I. A. & Norman, J. E. 2002. Immunolocalization of proinflammatory cytokines in myometrium, cervix, and fetal membranes during human parturition at term. *Biol Reprod,* 66**,** 445-9.

Yuan, S. Q., Cohen, D. B., Ravel, J., Abdo, Z. & Forney, L. J. 2012. Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome. *Plos One,* 7.

Zhou, X., Brown, C. J., Abdo, Z., Davis, C. C., Hansmann, M. A., Joyce, P., Foster, J. A. & Forney, L. J. 2007. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. *Isme j,* 1**,** 121-33.

Appendix

Table A 1 - Primers tested for use in qPCR analyses. Tests carried out at 20pmol/µl concentration. Sequences run within RDP probe match to identify number of potential matches to curated species within RDP database.

Primer pair	Primer Sequence (position on 16S rRNA *E. coli* sequence)	Length (bp)	N RDP matches	PCR efficiency	Reference
785F-1175R	GGATTAGATACCCBRGTAGTC - ACGTCRTCCCCDCCTTCCTC	390	1326070	53%	Doyle et al. (2014)
785F-907R	GGATTAGATACCCBRGTAGTC - CCGTCAATTCMTTTRAGTTT	122	1547022	140%	Morales and Holben (2009)
U337F-U518R	GACTCCTACGGGAGGCWGCAG - GTATTACCGCGGCTGCTGG	181	1464742	115%	Jaric et al. (2013)
785F-984R	GGATTAGATACCCBRGTAGTC - GTAAGGTTCYTCGCGT	199	1266536	81%	Wang and Qian (2009)
785F-939R	GGATTAGATACCCBRGTAGTC - CTTGTGCGGGYCCCCGTCAAT	154	1473011	84%	Wang and Qian (2009)
785F-926R	GGATTAGATACCCBRGTAGTC - CCGTCAATTCMTTTGAGT	141	1550814	81%	Wang and Qian (2009)
785F-E939R	GGATTAGATACCCBRGTAGTC - CTTGTGCGGGCCCCCGTCAATTC	154	1024705	86%	Wang and Qian (2009)

Table A 2 – Summary characteristics of qPCR plates run for analysis of bacterial load in placental tissue. Concentrations quantified using PCR, slope R2 and efficiency refer to plate specific standard curves. Three negative controls were run per plate to check for bacterial contamination in PCR reagents.

Plate	Concentration of standard (pg/ul)	Slope	R^2	Efficiency (%)	Negative control 1 (CT)	Negative control 2 (CT)	Negative control 3 (CT)
1	810	-3.90	0.99	80.16	33.10	31.47	NA
2	810	-3.66	0.99	87.51	32.48	32.64	33.72
3	810	-3.57	0.99	90.54	33.36	35.53	32.13
4	810	-3.57	0.99	90.71	32.61	31.93	31.62
5	810	-3.30	0.99	100.83	30.81	29.85	31.83
6	550	-3.57	0.96	90.75	33.01	32.46	32.85
7	550	-3.51	0.99	92.87	32.91	32.69	33.55
8	550	-3.58	0.96	90.36	31.95	33.34	33.00
9	550	-3.76	0.98	84.52	33.46	32.44	33.45
10	550	-3.53	0.97	92.06	32.84	33.59	32.42
11	550	-3.58	0.98	90.42	31.95	32.13	32.30
12	550	-3.49	0.98	93.60	32.18	32.99	32.74
13	376	-3.53	0.97	92.08	34.69	33.51	33.31
14	376	-3.31	0.97	100.50	34.32	35.01	34.59
15	376	-3.48	0.99	93.70	32.57	32.93	32.78
16	376	-3.52	0.99	92.24	32.92	33.01	NA
17	376	-3.33	0.99	99.52	30.81	31.34	31.10
18	376	-3.40	0.96	96.80	32.36	34.97	33.21
19	376	-3.58	0.97	90.42	32.72	32.83	34.52
20	376	-3.44	0.96	95.33	30.60	33.27	33.45
21	412	-3.34	0.96	99.25	31.77	32.91	32.77
22	412	-3.31	0.96	100.41	32.10	31.97	29.94
23	412	-3.24	0.97	103.34	31.25	31.34	31.65
24	412	-3.43	0.98	95.86	29.89	32.86	32.62
25	412	-3.66	1.00	87.51	32.72	31.85	29.96
26	412	-3.69	0.98	86.63	31.36	31.60	28.42
27	412	-3.32	0.98	100.16	29.97	32.67	32.10
28	412	-3.37	0.99	98.15	32.54	31.94	32.66
29	412	-3.31	0.97	100.66	32.35	32.77	32.99
30	412	-3.59	0.99	89.79	31.39	31.87	27.60
31	412	-3.40	0.99	96.74	32.78	33.16	28.35
32	628	-3.48	0.99	93.84	34.37	34.56	33.46
33	628	-3.62	0.99	89.01	32.68	35.18	31.64
34	628	-3.50	0.98	93.12	32.38	34.51	31.68
35	628	-3.20	0.97	105.47	32.73	32.89	32.46

Table A 3 – List of potential contaminating genera present in extraction reagents. Contaminants defined as any OTUs with at least two reads in at least two of negative extraction samples. Red = all OTUs mapping to genus removed from experimental samples; Orange = OTUs present in negative extracts removed from samples; Green = not removed from samples.

Genus	N OTUs in negative extractions	N OTUs in placental samples	Mean % abundance negative extractions	Mean % abundance placental samples	Listed in Glassing et al. (2016)?
Acidovorax	2	2	1.67	0.06	Yes
Aeromonas	1	1	0.05	0.04	No
Afipia	1	1	0.09	0.04	Yes
Arabidopsis	2	2	0.41	0.08	No
Aureimonas	1	1	0.22	0	No
Brevibacterium	3	21	0.23	0.05	Yes
Chryseobacterium	1	56	0.45	0.05	Yes
Cloacibacterium	1	9	0.03	0.38	Yes
Enhydrobacter	3	15	5.39	0.89	Yes
Janibacter	1	7	0.05	0.12	Yes
Klebsiella	1	1	0.11	0.02	Yes
Massilia	6	6	11.68	0.8	Yes
Paracoccus	1	17	0.02	0.01	Yes
Polynucleobacter	1	2	0.05	0	No
Ralstonia	2	3	0.26	0.03	Yes
Rheinheimera	1	6	0.02	0	No
Variovorax	2	2	0.65	0.05	Yes
Acinetobacter	25	78	20.44	4.54	Yes
Bacillus	1	85	0	0	Yes
Corynebacterium	4	117	0.93	0.58	Yes
Delftia	2	11	1.12	0.19	Yes
Dermacoccus	1	7	0.88	0.01	No
Enterococcus	1	89	0.24	0.17	Yes
Escherichia	6	39	4.7	1.5	Yes
Finegoldia	1	6	0.08	0.06	No
Granulicatella	1	9	0.09	0.07	Yes
Haemophilus	1	20	0.04	0.02	Yes
Kocuria	7	17	1.7	0.12	Yes
Lactococcus	1	6	2.3	0.06	No
Micrococcus	1	9	1.55	0.58	Yes
Mycoplana	1	13	0.07	0.3	No
Neisseria	2	18	0.96	0.05	Yes
Peptoniphilus	1	20	0.19	0.1	Yes
Peptostreptococcus	1	89	0.01	0.05	No
Propionibacterium	6	47	4.28	1.34	Yes
Pseudomonas	4	75	4.04	1.3	Yes
Rothia	3	28	0.74	0.14	Yes
Sphingomonas	3	25	1.31	0.63	Yes
Staphylococcus	15	205	8.14	3.5	Yes
Streptococcus	7	390	2.63	2.53	Yes
Unassigned	5	2813	2.25	0.7	No
Lactobacillus	2	120	0.27	17.89	Yes
Mycoplasma	1	19	0.01	1.66	No
Veillonella	1	22	0.22	8.18	Yes

Table A 4 – Distributions of types of missingness in optimisation experiments by cytokine and batch.

Cytokine	IL-1β		IL1ra		IL2		IL4		IL5		IL6		IL7		IL8		IL9	
Batch	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Below blank	0	0	0	3	16	18	0	0	6	5	0	0	0	0	0	0	1	0
OOR below	0	0	3	6	7	7	3	0	8	4	0	0	4	6	0	0	13	2
OOR above	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Extrapolated beyond curve	5	3	0	0	0	1	0	0	3	9	0	0	0	0	0	0	0	0
%CV >20	0	1	0	1	0	1	0	1	0	1	0	1	3	3	0	1	4	3
Total N (%)	5 (13)	3 (8)	3 (8)	9 (25)	23 (63)	26 (72)	3 (8)	0 (0)	17 (47)	18 (50)	0 (0)	0 (0)	4 (11)	6 (16)	0 (0)	0 (0)	14 (38)	2 (5)

Cytokine	IL10		IL12		IL13		IL15		IL17A		CCL11		FGF basic		G-CSF		GM-CSF	
Batch	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Below blank	8	3	8	9	3	6	17	25	10	12	0	0	9	7	0	0	33	28
OOR below	0	0	0	0	7	2	6	5	4	0	0	0	5	1	9	0	1	2
OOR above	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Extrapolated beyond curve	0	6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
%CV >20	0	1	0	1	3	4	0	1	0	1	0	1	0	1	2	1	1	0
Total N (%)	8 (22)	9 (19)	8 (22)	10 (27)	10 (27)	8 (22)	23 (63)	30 (83)	14 (38)	12 (33)	0 (0)	0 (0)	14 (38)	8 (22)	9 (25)	0 (0)	34 (94)	30 (83)

Cytokine	IFN-γ		CXCL10		CCL2		CCL3		PDGFbb		CCL4		RANTES		TNF-α		VEGF	
Batch	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Below blank	0	0	0	0	14	2	3	0	0	0	0	0	0	0	8	3	18	14
OOR below	36	19	0	0	5	10	0	0	0	0	0	0	0	0	0	0	7	0
OOR above	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0
Extrapolated beyond curve	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	1	3
%CV >20	0	1	0	0	0	3	0	1	2	2	0	0	0	0	0	1	0	1
Total N (%)	36 (100)	19 (52)	0 (0)	0 (0)	19 (52)	12 (33)	3 (8)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	31 (86)	0 (0)	9 (25)	7 (19)	26 (72)	17 (47)

Table A 5 – Summary of cytokine concentration distributions when samples are diluted 1:2 or 1:4. Comparisons of medians conducted using Wilcox test.

Cytokine	1:2 Dilution		1:4 Dilution		Difference between medians	Difference between ranges	W	P
	Median	Min-Max (Range)	Median	Min-Max (Range)				
IL-1β	2.52	0.34-6.44 (6.1)	3.99	0.19-8.21 (8.02)	1	1.92	527	0.175
IL-1Ra	146.24	0-399.77 (399.77)	223.17	0-517.87 (517.87)	77	118.1	519	0.148
IL2	2.60	0.125-26.39 (26.27)	2.60	0.125-38.62 (38.50)	0	12.23	627	0.806
IL4	5.54	0.64-15.33 (14.69)	8.03	0.15-25.6 (25.45)	2	10.76	458	0.033
IL5	4.20	0.03-16.49 (16.46)	5.08	0.03-23.77 (23.74)	1	7.28	570	0.489
IL6	17.10	0-27.27 (27.27)	23.36	3.07-36.56 (33.49)	6	6.22	327	0.00031
IL7	6.75	0-13.71 (13.71)	7.33	0-34.65 (34.65)	1	20.94	486	0.353
IL8	21.51	0-53.14 (53.14)	37.73	0-83.83 (83.83)	16	30.69	424	0.012
IL9	11.15	0-31.89 (31.89)	26.90	0-47.32 (47.32)	16	15.43	341	0.004
IL10	0.62	0-21.38 (21.38)	0.00	0-32.49 (32.49)	-1	11.11	462	0.053
IL12	11.57	0-34.88 (34.88)	28.36	0-73.58 (73.58)	17	38.7	444	0.021
IL13	5.05	0.225-13.12 (12.90)	5.49	0.225-12.4 (12.175)	0	-0.72	502	0.594
IL15	3.33	0-24.09 (24.09)	0.68	0-33.58 (33.58)	-3	9.49	706	0.469
IL17A	33.12	0.85-124.14 (123.29)	99.50	0.85-309.61 (308.76)	66	185.47	464	0.036
CCL11	0.00	0-49.74 (49.74)	0.00	0-70.08 (70.08)	0	20.34	354	0.0010
FGF basic	15.07	0.465-98.17 (97.71)	46.92	0.465-236 (235.54)	32	137.83	467	0.059
G-CSF	60.31	6.13-158.52 (152.39)	99.75	6.13-233.61 (227.48)	39	75.09	407	0.016
GM-CSF	1.55	0-5.96 (5.96)	1.55	0-71.88 (71.88)	0	65.92	501	0.006
IFN-γ	66.05	43.865-178.65 (134.79)	66.05	0-250.55 (250.55)	0	115.765	610	0.64
CXCL10	280.37	90.24-562.73 (472.49)	353.00	191.35-710.74 (519.39)	73	46.9	427	0.012
CCL2	17.91	3.61-55.18 (51.57)	28.98	3.61-105.28 (101.67)	11	50.1	550	0.457
CCL3	2.63	0.42-5.7 (5.28)	7.59	0.085-13.91 (13.83)	5	8.545	354	0.0010
PDGF bb	495.58	0-2688.37 (2688.37)	788.54	0-3548.13 (3548.13)	293	859.76	445	0.11
CCL4	0.00	0-54.03 (54.03)	37.67	0-96.19 (96.19)	38	42.16	201	9.9E-08
RANTES	1323.94	1064.36-6263.08 (5198.72)	2873.28	1323.94-15485.9 (14161.96)	1549	8963.24	421	0.0080
TNF-α	25.27	0.415-84.68 (84.27)	42.52	0.42-120.41 (119.99)	17	35.725	498	0.092
VEGF	0.31	0.27-27.17 (26.9)	1.26	0.27-53.66 (53.39)	1	26.49	465	0.031

Table A 6 – Adjusted and unadjusted regression models for impact of maternal BMI on GA at birth (weeks).

Predictors	Unadjusted			Adjusted		
	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	38.45 (38.32-38.59)	0.069	<2E-16	38.72 (38.56-38.88)	0.08	< 2E-16
Maternal BMI (baseline = Normal)						
Obese	-0.32 (-0.67-0.04)	0.18	0.080	-0.16 (-0.54-0.21)	0.19	3.83E-01
Underweight	-0.25 (-0.98-0.48)	0.37	0.50	-0.27 (-1.03-0.49)	0.39	4.84E-01
Maternal ethnicity (baseline =White, white British)						
Asian, Asian British				-1.36 (-1.96--0.75)	0.31	1.30E-05
Black, Black British				-1.22 (-1.66--0.79)	0.22	5.44E-08
All others				-0.57 (-0.9--0.23)	0.17	1.11E-03

Table A 7 – Adjusted and unadjusted regression models for impact of maternal BMI on odds of having a term versus preterm birth.

Predictors	Unadjusted			Adjusted		
	OR (95% CI)	β SE	P	OR (95% CI)	β SE	P
Intercept	6.03 (5.26-6.95)	0.071	<2E-16	7.64 (6.4-9.2)	0.092	< 2E-16
Maternal BMI (baseline=Normal)						
Obese	0.78 (0.56-1.1)	0.17	0.15	0.86 (0.61-1.25)	0.18	0.43
Underweight	0.69 (0.37-1.43)	0.34	0.29	0.63 (0.33-1.31)	0.35	0.19
Maternal ethnicity (baseline = White, white British)						
Asian, Asian British				0.47 (0.28-0.83)	0.28	0.0063
Black, Black British				0.39 (0.27-0.58)	0.20	2.08E-06
All others				0.66 (0.47-0.93)	0.17	0.017

Table A 8 – Adjusted and unadjusted regression models for impact of maternal diabetes status on GA at birth (weeks).

Predictors	Unadjusted			Adjusted		
	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	38.29 (38.14-38.44)	0.08	< 2E-16	38.59 (38.39-38.78)	0.10	< 2E-16
Maternal diabetes (baseline=None)						
Gestational	-0.87 (-1.62--0.12)	0.38	0.023	-0.75 (-1.53-0.03)	0.40	0.061
Established	-1.38 (-2.36--0.39)	0.50	0.0063	-1.16 (-2.17--0.14)	0.52	0.025
Maternal ethnicity (baseline= White, white British)						
Asian, Asian British				-1.86 (-2.61--1.11)	0.38	1.27E-06
Black, Black British				-1.21 (-1.69--0.73)	0.25	9.35E-07
All others				-0.52 (-0.93--0.11)	0.21	0.013

Table A 9 – Adjusted and unadjusted regression models for showing impact of maternal diabetes status on odds of having a term over preterm birth. The odds of having a preterm birth in mothers with gestational or established diabetes are higher than those of having a term birth. This remains significant for gestational but not established diabetes status, when adjusted for the confounding effects of maternal ethnicity.

Predictors	Unadjusted			Adjusted		
	OR (95% CI)	β SE	P	OR (95% CI)	β SE	P
(Intercept)	5.27 (4.58-6.09)	0.07	<2E-16	6.6 (5.46-8.06)	0.10	< 2E-16
Maternal diabetes (baseline=None)						
Gestational	0.44 (0.25-0.8)	0.29	0.0051	0.45 (0.25-0.84)	0.31	0.01
Established	0.46 (0.22-1.01)	0.38	0.0402	0.61 (0.28-1.48)	0.42	0.24
Maternal ethnicity (baseline= White, white British)						
Asian, Asian British				0.38 (0.21-0.69)	0.30	0.00103
Black, Black British				0.43 (0.29-0.64)	0.20	2.63E-05
All others				0.71 (0.49-1.05)	0.19	0.08

Table A 10 - Adjusted and unadjusted models for effect of a one-week increase in GA at birth on the odds of having a vaginal over a CS delivery. The odds of having a vaginal rather than a CS delivery increase with every one-week increase in GA at birth. This relationship remains significant with adjustment for potential confounders.

Predictor	Unadjusted			Adjusted		
	OR (95% CI)	β SE	P	OR (95% CI)	β SE	P
Intercept	0.0016 (0.00030-0.0077)	0.83	6.06E-15	0.0081 (0.0013-0.048)	0.92	1.76E-07
GA at birth	**1.20 (1.15-1.25)**	**0.022**	**<2e-16**	**1.21 (1.16-1.26)**	**0.022**	**<2E-16**
Parity (baseline=0)						
1				0.95 (0.74-1.21)	0.12	0.68
2 or 3				1.36 (0.97-1.92)	0.17	0.074
4 or more				1.91 (0.77-5.08)	0.48	0.17
Maternal age at booking				0.94 (0.92-0.96)	0.011	3.87E-08
Maternal diabetes (baseline=No)						
Established				0.31 (0.58-1.73)	0.28	0.0294
Gestational				1.00 (0.09-0.81)	0.54	0.99
Maternal BMI						0.025
Obese				0.68 (0.49-0.93)	0.16	0.015
Underweight				1.18 (0.65-2.26)	0.32	0.596

Table A 11 – Summary table showing comparison in distributions of key clinical characteristics between preterm pregnancies within the BBB taken forward for qPCR analyses of placental tissue, compared to those without available placental tissue.

Clinical characteristic	Sub-group	qPCR preterm cohort (N=146)	BBB preterm samples without placental tissue (N=145)	Wilcox or X2
GA at birth		Mean = 33.5 SD = 3.01 Median = 35 IQR = 33-36	Mean = 32.8 SD = 3.8 Median = 34 IQR = 31-36	W=11354 P=0.27
Birthweight		Mean = 2122 SD =745 Median = 2230 IQR = 1605-2600	Mean = 2012 SD = 798.2 Median = 2080 IQR = 1510-2540	W=10755 P=0.23
Delivery method				
	Vaginal	54 (37%)	47 (32.4%)	$X^2 = 0.00026$ P=0.99
	CS	87 (59.6%)	73 (50.3%)	
Labour onset				
	No labour	88 (60.3%)	74 (51.0%)	$X^2 = 0.90$ P=0.34
	Spontaneous	44 (30.1%)	49 (34.0%)	
Membrane rupture				
	None	81 (55.5%)	68 (46.9%)	$X^2 = 0.73$ P = 0.39
	Spontaneous	51 (34.9%)	55 (37.9%)	

Table A 12 - Adjusted multi-level linear regression model for association between bacterial load (log 16S copy number) and pregnancy outcome in whole cohort.

Predictor		β (95% CI)	β SE	P
Intercept		4.67 (4.25-5.09)	0.21	<2.2E-16
Outcome	**nsPTB**	**-0.29 (-0.5--0.08)**	**0.11**	**0.0064**
(baseline=sPTB)	**Term**	**-0.13 (-0.3-0.04)**	**0.09**	**0.13**
Delivery method	Vaginal	-0.02 (-0.15-0.12)	0.07	0.79
	Asian	0.05 (-0.22-0.31)	0.14	0.73
Maternal ethnicity	Black	0.29 (0.09-0.48)	0.10	0.0047
	All others	-0.05 (-0.22-0.12)	0.09	0.58
	2	1.33 (0.93-1.73)	0.20	4.3E-10
	3	0.09 (-0.36-0.55)	0.23	0.69
	4	0.63 (0.21-1.05)	0.21	3.4E-03
	5	0.17 (-0.44-0.77)	0.31	0.59
	6	0.47 (0.03-0.91)	0.22	0.037
	7	0.03 (-0.36-0.43)	0.20	0.87
	8	0.5 (0.07-0.93)	0.22	0.024
	9	1.48 (1.06-1.91)	0.21	4.2E-11
	10	0.72 (0.29-1.14)	0.21	9.8E-04
	11	1.19 (0.81-1.57)	0.19	2.6E-09
	12	0.93 (0.53-1.32)	0.20	5.5E-06
	13	0.55 (0.17-0.94)	0.20	0.0054
	14	0.15 (-0.24-0.54)	0.20	0.46
	15	-0.09 (-0.52-0.33)	0.22	0.67
	16	-0.01 (-0.44-0.41)	0.22	0.95
	17	0.04 (-0.45-0.52)	0.25	0.88
Plate	18	0.1 (-0.36-0.57)	0.24	0.66
	19	0.44 (0.03-0.86)	0.21	0.035
	20	0.7 (0.22-1.19)	0.25	0.0046
	21	-0.35 (-0.76-0.06)	0.21	0.097
	22	-0.39 (-0.81-0.03)	0.21	0.071
	23	0.15 (-0.24-0.55)	0.20	0.44
	24	0.35 (-0.07-0.76)	0.21	0.10
	25	0.85 (0.44-1.27)	0.21	6.60E-05
	26	0.65 (0.03-1.26)	0.31	0.039
	27	0.28 (-0.15-0.72)	0.22	0.20
	28	0.05 (-0.39-0.48)	0.22	0.84
	29	-0.54 (-0.98--0.09)	0.22	0.018
	30	0.38 (-0.06-0.82)	0.22	0.089
	31	-0.31 (-0.76-0.13)	0.23	0.17
	32	0.49 (0.03-0.95)	0.23	0.037
	33	0.68 (0.27-1.09)	0.21	1.20E-03
	34	0.33 (-0.05-0.71)	0.19	0.084
	35	-0.43 (-0.85-0)	0.22	0.048
Smoking	No	-0.09 (-0.33-0.15)	0.12	0.46
Maternal BMI	Obese	-0.02 (-0.18-0.15)	0.08	0.86
	Underweight	-0.14 (-0.44-0.16)	0.15	0.36
Recruitment centre	QCCH	-0.05 (-0.31-0.2)	0.13	0.69
	SMH	-0.01 (-0.14-0.12)	0.07	0.86
Tissue type	Villous	0.07 (-0.11-0.25)	0.09	0.43

Table A 13 - Adjusted multi-level linear regression model for association between bacterial load (log 16S copy number) and pregnancy outcome in parenchyma tissue.

Predictor		β (95% CI)	β SE	P
Intercept		4.74 (4.32-5.16)	0.21	<2.2E-16
Outcome (baseline=sPTB)	**nsPTB**	**-0.33 (-0.53--0.12)**	**0.11**	**0.0022**
	Term	**-0.15 (-0.32-0.02)**	**0.09**	**0.076**
Delivery method	Vaginal	-0.05 (-0.19-0.08)	0.07	0.43
Ethnicity	Asian	0.07 (-0.2-0.33)	0.13	0.63
	Black	0.25 (0.05-0.44)	0.10	0.013
	All others	-0.02 (-0.2-0.15)	0.09	0.79
Plate	2	1.31 (0.92-1.7)	0.20	4.93E-10
	3	0.14 (-0.3-0.58)	0.22	0.53
	4	0.64 (0.23-1.05)	0.21	0.0025
	5	0.06 (-0.53-0.66)	0.30	0.84
	6	0.46 (0.02-0.89)	0.22	0.039
	7	0.04 (-0.34-0.43)	0.19	0.82
	8	0.53 (0.11-0.95)	0.21	0.015
	9	1.49 (1.08-1.91)	0.21	3.27E-11
	10	0.73 (0.32-1.14)	0.21	0.00057
	11	1.19 (0.82-1.56)	0.19	1.9E-09
	12	0.92 (0.54-1.3)	0.19	4.3E-06
	13	0.55 (0.17-0.93)	0.19	0.0046
	14	0.17 (-0.21-0.55)	0.19	0.38
	15	-0.12 (-0.53-0.3)	0.21	0.58
	16	-0.02 (-0.43-0.4)	0.21	0.94
	17	0.05 (-0.47-0.56)	0.26	0.86
	19	0.28 (-0.19-0.75)	0.24	0.25
	20	0.69 (-0.17-1.54)	0.43	0.12
	21	-0.31 (-0.71-0.09)	0.20	0.13
	22	-0.35 (-0.78-0.07)	0.22	0.10
	23	0.16 (-0.23-0.55)	0.20	0.41
	24	0.23 (-0.19-0.65)	0.21	0.28
	25	0.77 (0.36-1.18)	0.21	0.00030
	26	0.67 (0.08-1.27)	0.30	0.027
	27	0.39 (-0.09-0.87)	0.24	0.12
	28	0.1 (-0.34-0.55)	0.23	0.64
	29	-0.58 (-1.08--0.07)	0.26	0.025
	30	0.45 (-0.03-0.93)	0.24	0.065
	31	-0.24 (-0.71-0.22)	0.23	0.30
	32	0.54 (0.08-1)	0.23	0.022
	33	0.8 (0.36-1.23)	0.22	0.00037
	34	0.34 (-0.05-0.72)	0.20	0.090
	35	-0.29 (-0.73-0.15)	0.22	0.20
Smoking	No	-0.15 (-0.4-0.1)	0.13	0.24
Maternal BMI	Obese	0 (-0.16-0.16)	0.08	0.98
	Underweight	-0.11 (-0.42-0.19)	0.16	0.47
Recruitment hospital	QCCH	-0.01 (-0.28-0.25)	0.14	0.93
	SMH	0 (-0.12-0.13)	0.06	0.98

Table A 14 - Adjusted multi-level linear regression model for association between bacterial load (log 16S copy number) and pregnancy outcome in villous tissue.

Predictor		β (95% CI)	β SE	P
Intercept		5.25 (3.72-6.78)	0.76	7.04E-09
Outcome (baseline=sPTB)	**nsPTB**	**-0.59 (-1.43-0.26)**	**0.42**	**0.17**
	Term	**-0.32 (-1.01-0.36)**	**0.34**	**0.35**
Delivery method	Vaginal	0.09 (-0.4-0.57)	0.24	0.72
Maternal ethnicity	Asian, Asian British	0.46 (-0.6-1.52)	0.53	0.39
	Black, Black British	0.71 (-0.15-1.57)	0.43	0.10
	All others	-0.26 (-0.82-0.31)	0.28	0.37
Plate	18	-0.07 (-1.43-1.29)	0.68	0.92
	19	0.35 (-1.04-1.73)	0.69	0.62
	20	0.41 (-0.95-1.76)	0.68	0.55
	22	-0.72 (-2.51-1.07)	0.89	0.42
	23	-0.19 (-1.99-1.62)	0.90	0.84
	24	0.78 (-0.89-2.46)	0.84	0.35
	25	1.17 (-0.64-2.99)	0.91	0.20
	27	-0.46 (-1.96-1.04)	0.75	0.54
	28	-0.92 (-2.62-0.77)	0.85	0.28
	29	-0.95 (-2.38-0.48)	0.71	0.19
	30	0.21 (-1.26-1.69)	0.73	0.77
	31	-0.76 (-2.33-0.8)	0.78	0.33
	32	-0.67 (-2.83-1.48)	1.08	0.53
	33	0.03 (-1.44-1.5)	0.73	0.96
	34	-0.03 (-1.55-1.5)	0.76	0.97
	35	-1.02 (-2.52-0.48)	0.75	0.18
Smoking	No	-0.07 (-0.75-0.6)	0.34	0.83
Maternal BMI	Obese	-0.17 (-0.91-0.58)	0.37	0.66
	Underweight	-0.68 (-1.66-0.3)	0.49	0.17
Recruitment hospital	QCCH	-0.15 (-0.85-0.56)	0.35	0.68
	SMH	-0.72 (-1.69-0.25)	0.48	0.14

Table A 15 - Adjusted regression model for association between placental bacterial load (log 16S copy number) and cervical suture in current pregnancy

Predictor		β (95% CI)	β SE	P
Intercept		4.65 (4.21-5.09)	0.22	6.94E-58
Suture	**No**	**-0.31 (-0.64-0.02)**	**0.17**	**0.068**
Delivery	Vaginal	0.01 (-0.13-0.16)	0.07	0.84
	2	1.53 (1.12-1.94)	0.21	4.9E-12
	3	0.19 (-0.28-0.66)	0.24	0.43
	4	0.86 (0.43-1.29)	0.22	0.00012
	5	0.39 (-0.27-1.05)	0.33	0.24
	6	0.74 (0.24-1.23)	0.25	0.0038
	7	0.28 (-0.13-0.7)	0.21	0.18
	8	0.53 (0.1-0.97)	0.22	0.016
	9	1.69 (1.24-2.13)	0.23	3.5E-12
	10	0.87 (0.43-1.31)	0.22	0.00014
	11	1.42 (1.05-1.8)	0.19	3.7E-12
	12	1.11 (0.71-1.51)	0.20	1.3E-07
	13	0.73 (0.33-1.12)	0.20	0.00036
	14	0.26 (-0.15-0.66)	0.21	0.21
	15	0.14 (-0.31-0.58)	0.23	0.54
	16	0.25 (-0.25-0.76)	0.26	0.33
	17	-0.24 (-1.49-1.01)	0.63	0.71
Plate	18	0.18 (-0.32-0.68)	0.25	0.48
	19	0.55 (0.11-0.98)	0.22	0.014
	20	0.89 (0.37-1.41)	0.26	0.00082
	21	-0.14 (-0.56-0.27)	0.21	0.50
	22	-0.17 (-0.62-0.27)	0.22	0.44
	23	0.36 (-0.04-0.77)	0.21	0.078
	24	0.55 (0.13-0.98)	0.21	0.011
	25	1.06 (0.63-1.49)	0.22	2.5E-06
	26	0.85 (0.24-1.46)	0.31	0.0069
	27	0.52 (0.07-0.96)	0.23	0.023
	28	0.2 (-0.25-0.65)	0.23	0.39
	29	-0.3 (-0.8-0.2)	0.25	0.24
	30	0.55 (0.06-1.03)	0.25	0.028
	31	-0.2 (-0.67-0.28)	0.24	0.42
	32	0.66 (0.19-1.14)	0.24	0.0063
	33	0.89 (0.47-1.31)	0.21	4.12E-05
	34	0.59 (0.19-0.98)	0.20	0.0039
	35	-0.29 (-0.72-0.14)	0.22	0.18
Outcome	nsPTB	-0.19 (-0.41-0.03)	0.11	0.098
	Term	-0.11 (-0.28-0.07)	0.09	0.24
Tissue	Villous	0.05 (-0.14-0.25)	0.10	0.59

Table A 16 – Log2 fold change of differentially abundant OTUs between vaginally and CS delivered placenta. Q values are Benjamini-Hochberg adjusted P values.

OTU	Genus	Log2 fold change	P	Q
519673	*Lactobacillus*	-3.15 (-4.16--2.14)	2.96E-09	3.65E-06
823803	*Lactobacillus*	-1.15 (-1.57--0.72)	2.18E-07	1.35E-04
820367	*Lactobacillus*	-0.68 (-0.94--0.41)	1.03E-06	3.39E-04
593376	*Lactobacillus*	-0.79 (-1.1--0.48)	1.10E-06	3.39E-04
888575	*Lactobacillus*	-1.36 (-1.94--0.79)	5.28E-06	1.22E-03
813785	*Lactobacillus*	-1.15 (-1.64--0.66)	5.92E-06	1.22E-03
553352	*Lactobacillus*	-0.51 (-0.74--0.28)	2.05E-05	3.62E-03
NROTU60	*Lactobacillus*	-0.93 (-1.36--0.5)	2.95E-05	4.56E-03
815380	*Lactobacillus*	-0.51 (-0.75--0.27)	3.56E-05	4.88E-03
806179	*Lactobacillus*	-0.44 (-0.66--0.23)	6.45E-05	7.97E-03
4447432	*Lactobacillus*	-0.82 (-1.23--0.41)	1.00E-04	1.12E-02
817903	*Lactobacillus*	-0.55 (-0.83--0.26)	1.92E-04	1.92E-02
498355	*Lactobacillus*	-0.55 (-0.83--0.26)	2.01E-04	1.92E-02
354225	*Lactobacillus*	-0.58 (-0.89--0.27)	2.63E-04	2.32E-02
316515	*Lactobacillus*	-0.3 (-0.47--0.13)	5.41E-04	4.18E-02
335967	*Lactobacillus*	-0.26 (-0.4--0.11)	5.41E-04	4.18E-02
255367	*Lactobacillus*	-0.35 (-0.55--0.14)	9.02E-04	6.56E-02
308463	*Lactobacillus*	-0.26 (-0.41--0.1)	1.20E-03	8.22E-02
350349	*Lactobacillus*	-0.32 (-0.52--0.13)	1.26E-03	8.22E-02
NROTU138	*Lactobacillus*	-0.65 (-1.05--0.26)	1.35E-03	8.32E-02
840914	*Prevotella*	-0.62 (-1--0.24)	1.47E-03	8.59E-02
956654	*Lactobacillus*	-1.62 (-2.61--0.62)	1.53E-03	8.59E-02
366068	*Faecalibacterium*	-0.47 (-0.77--0.18)	1.63E-03	8.75E-02
132873	*Lactobacillus*	-0.35 (-0.57--0.13)	2.26E-03	1.16E-01
365717	*Faecalibacterium*	-0.38 (-0.63--0.13)	3.05E-03	1.49E-01
622256	*Ruminococcus*	-0.49 (-0.82--0.17)	3.12E-03	1.49E-01
697479	*Tepidimonas*	1.1 (0.35-1.86)	4.19E-03	1.89E-01
4200918	*Lactobacillus*	-0.3 (-0.5--0.09)	4.27E-03	1.89E-01
4348210	*Lactobacillus*	-0.23 (-0.38--0.07)	4.44E-03	1.89E-01
583656	*Bacteroides*	-0.38 (-0.64--0.12)	4.72E-03	1.94E-01
818672	*Lactobacillus*	-0.22 (-0.38--0.07)	5.18E-03	2.07E-01
936190	*Streptococcus*	0.14 (0.04-0.23)	5.65E-03	2.18E-01
344154	*Bacteroides*	-0.49 (-0.83--0.14)	5.82E-03	2.18E-01
244464	*Lactobacillus*	-0.2 (-0.34--0.06)	6.07E-03	2.20E-01
804653	*Lactobacillus*	-0.19 (-0.32--0.05)	6.22E-03	2.20E-01
1015143	*Streptococcus*	0.18 (0.05-0.32)	6.63E-03	2.28E-01
944266	*Staphylococcus*	0.3 (0.08-0.51)	7.22E-03	2.41E-01
NROTU158	*Streptococcus*	0.5 (0.14-0.87)	7.52E-03	2.45E-01
958496	*Lactobacillus*	-0.15 (-0.26--0.04)	8.09E-03	2.56E-01
766571	*Staphylococcus*	0.16 (0.04-0.27)	8.27E-03	2.56E-01
813944	*Lactobacillus*	-0.2 (-0.34--0.05)	8.50E-03	2.56E-01
560336	*Bacteroides*	-0.34 (-0.6--0.09)	8.84E-03	2.56E-01
359650	*Faecalibacterium*	-0.18 (-0.31--0.05)	9.05E-03	2.56E-01
525215	*Faecalibacterium*	-0.3 (-0.53--0.08)	9.13E-03	2.56E-01
132829	*Lactobacillus*	-0.15 (-0.27--0.04)	9.59E-03	2.60E-01
1028283	*Prevotella*	-0.53 (-0.94--0.13)	9.66E-03	2.60E-01

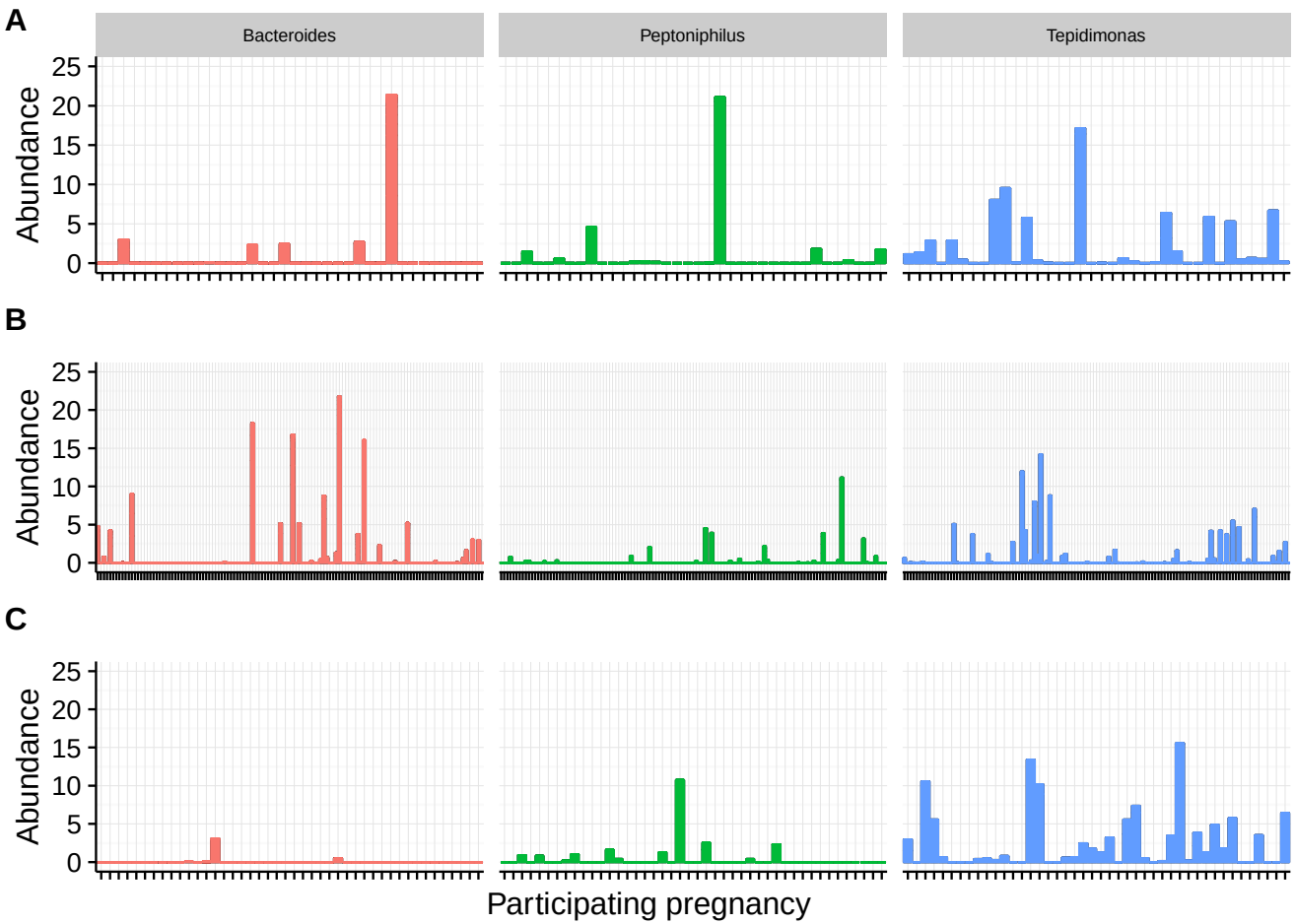


Figure A 1 - Abundance (%) of total reads in placentas from individual pregnancies by individual genera 1. A) sPTB placenta B) Term delivered placenta C) nsPTB placenta.

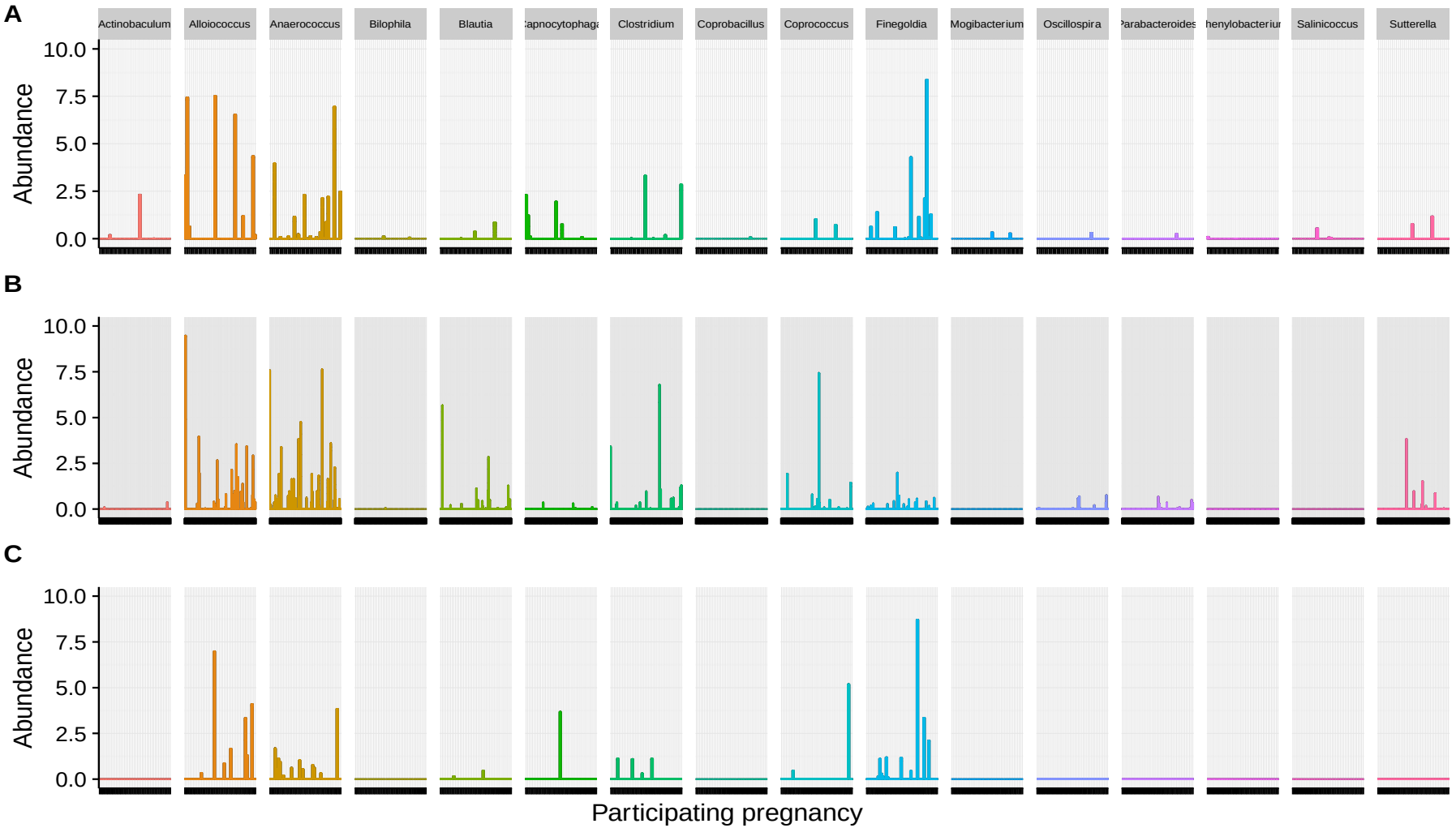


Figure A 2 - Abundance (%) of total reads in placentas from individual pregnancies by individual genera 2. A) sPTB placenta B) Term delivered placenta C) nsPTB placenta.

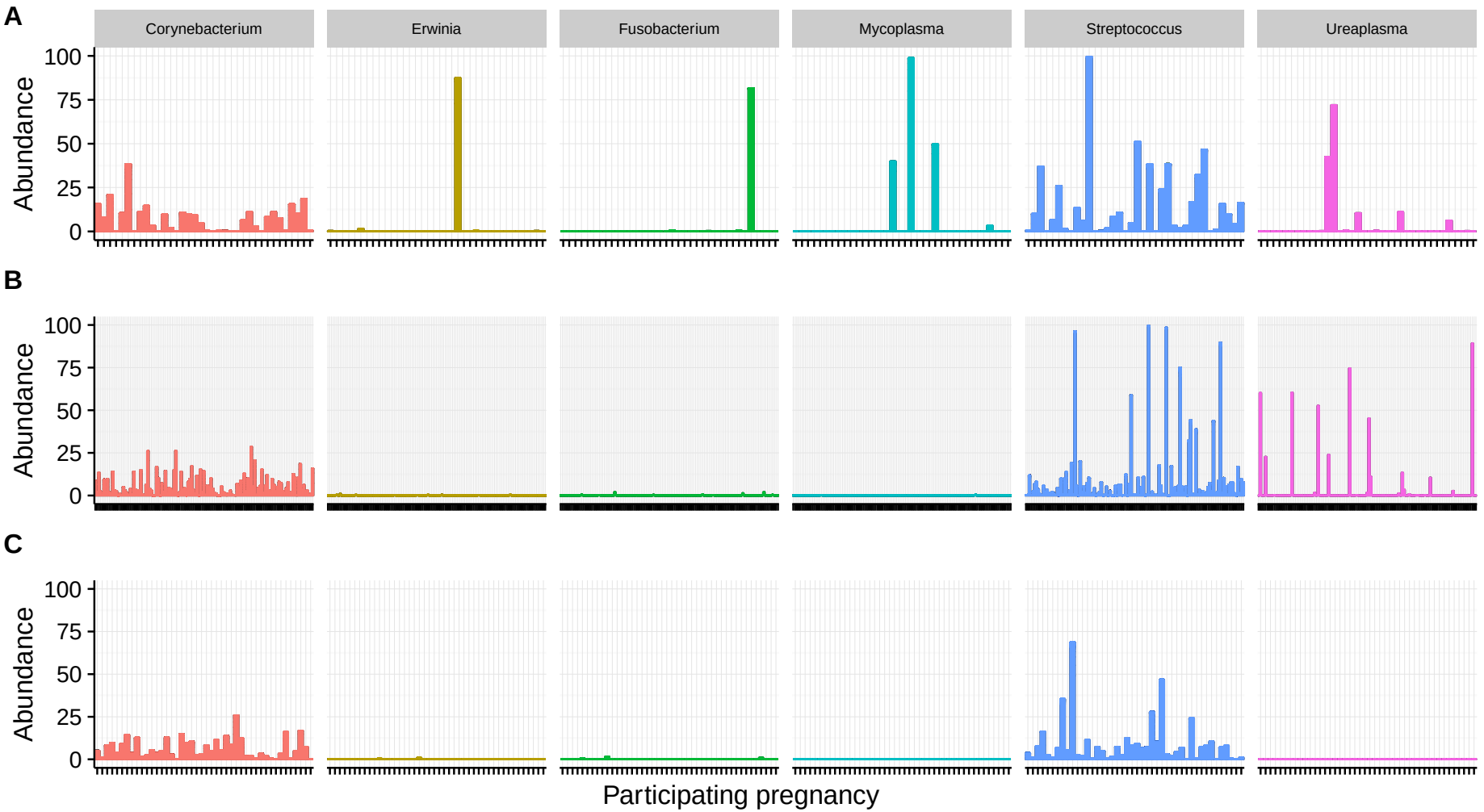


Figure A 3 - Abundance (%) of total reads in placentas from individual pregnancies by individual genera 3. A) sPTB placenta B) Term delivered placenta C) nsPTB placenta.

Table A 17 – Distribution of missingness (%) by cytokine and batch in 6 ELISA assays.

Batch	Type of missingness	IL-1β	IL1ra	IL2	IL4	IL5	IL6	IL7	IL8	IL9	IL10	IL12	IL13	IL15	IL17A	CCL11	FGF-basic	G-CSF	GM-CSF	IFN-γ	CXCL10	CCL2	CCL3	PDGF-bb	CCL4	RANTES	TNF-α	VEGF
Batch 1 (%)	LOD	0	0	47	0	0	0	0	0	0	12	0	0	82	0	0	0	0	65	0	0	65	0	0	0	0	0	12
	OOR	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	65	0	0
	Total	0	0	47	0	0	0	0	0	0	12	6	0	82	0	0	0	0	65	0	0	65	0	0	0	65	0	12
Batch 2 (%)	LOD	0	0	64	0	0	0	0	0	0	7	0	0	100	7	0	7	0	93	0	0	64	7	0	0	0	0	7
	OOR	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	71	0	0
	Total	0	0	64	0	7	0	0	0	0	7	0	0	100	7	0	7	0	93	0	0	64	7	0	0	71	0	7
Batch 3 (%)	LOD	0	0	25	0	0	0	0	0	0	6	0	0	75	0	0	0	0	62	0	0	56	0	0	0	0	0	6
	OOR	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	6	0	0	62	0	0
	Total	0	0	25	0	0	6	0	0	0	6	0	0	75	0	0	0	0	62	6	0	56	6	0	0	62	0	6
Batch 4 (%)	LOD	0	0	65	0	41	0	0	0	0	18	0	0	88	12	0	6	0	88	0	0	76	6	0	0	0	0	29
	OOR	6	6	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Total	6	6	65	0	41	6	0	0	0	18	0	0	88	12	0	6	0	88	0	0	76	6	0	0	0	0	29
Batch 5 (%)	LOD	0	0	74	0	34	0	0	0	0	0	3	47	76	0	0	0	0	42	0	0	39	0	0	0	0	0	3
	OOR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0
	Total	0	0	74	0	34	0	0	0	0	0	3	47	76	0	0	0	0	53	0	0	39	0	0	0	0	0	3
Batch 6 (%)	LOD	0	3	53	0	8	0	0	0	0	0	0	0	82	0	0	0	0	29	0	0	63	0	0	0	0	0	0
	OOR	0	36	0	0	0	0	0	100	0	0	0	0	13	3	0	0	0	16	0	0	0	0	0	0	0	0	0
	Total	0	39	53	0	8	0	0	100	0	0	0	0	95	3	0	0	0	45	0	0	63	0	0	0	0	0	0
All plates (%)	**LOD**	**0**	**1**	**57**	**0**	**16**	**0**	**0**	**0**	**0**	**5**	**1**	**13**	**82**	**2**	**0**	**1**	**0**	**54**	**0**	**0**	**58**	**1**	**0**	**0**	**0**	**0**	**7**
	OOR	**1**	**10**	**0**	**0**	**1**	**1**	**0**	**27**	**0**	**0**	**0**	**0**	**4**	**1**	**0**	**0**	**0**	**7**	**1**	**0**	**0**	**1**	**0**	**0**	**22**	**0**	**0**
	Total	**1**	**11**	**57**	**0**	**17**	**1**	**0**	**27**	**0**	**5**	**1**	**13**	**86**	**3**	**0**	**1**	**0**	**61**	**1**	**0**	**58**	**2**	**0**	**0**	**22**	**0**	**7**

Table A 18 – Intraclass correlation coefficients between concentrations individual cytokines of technical replicates in optimisation and experimental datasets. Grey=not taken forward to final analyses due to poor reproducibility.

Cytokine	ICC (95% CI)	
	Optimisation data (N=36)	Experimental data (N=14)
IL-1β	0.96 (0.92-0.98)	0.85 (0.6-0.95)
IL-1Ra	0.85 (0.72-0.92)	0.51 (0.02-0.82)
IL2	0.73 (0.53-0.85)	0.5 (0-0.81)
IL4	0.25 (-0.07-0.53)	-0.33 (-0.65-0.21)
IL5	0.88 (0.79-0.94)	0.42 (-0.09-0.78)
IL6	0.95 (0.91-0.98)	0.67 (0.26-0.89)
IL7	0.49 (0.2-0.7)	-0.2 (-0.58-0.35)
IL8	0.51 (0.23-0.72)	0.93 (0.81-0.98)
IL9	0.75 (0.57-0.87)	-0.15 (-0.55-0.39)
IL10	0.9 (0.81-0.95)	0.52 (0.04-0.83)
IL12	0.88 (0.78-0.94)	0.69 (0.29-0.9)
IL13	0.7 (0.48-0.83)	-0.07 (-0.5-0.46)
IL15	0.26 (-0.07-0.54)	0.33 (-0.18-0.73)
IL17A	0.93 (0.86-0.96)	0.63 (0.19-0.87)
CCL11	0.83 (0.7-0.91)	0.5 (0-0.81)
FGF-basic	0.84 (0.72-0.92)	0.39 (-0.12-0.76)
G-CSF	0.77 (0.6-0.88)	0.16 (-0.33-0.64)
GM-CSF	0.77 (0.6-0.88)	0.39 (-0.12-0.76)
IFN-γ	0.03 (-0.3-0.35)	0.12 (-0.37-0.61)
CXCL10	0.76 (0.59-0.87)	0.15 (-0.34-0.63)
CCL2	0.52 (0.24-0.72)	0.71 (0.31-0.9)
CCL3	0.85 (0.73-0.92)	0.79 (0.47-0.93)
PDGF-BB	0.72 (0.52-0.85)	0.02 (-0.44-0.54)
CCL4	0.88 (0.78-0.94)	0.83 (0.56-0.95)
RANTES	-0.52 (-0.72--0.24)	-0.06 (-0.5-0.47)
TNF-α	0.85 (0.72-0.92)	0.48 (-0.02-0.81)
VEGF	0.78 (0.62-0.88)	0.7 (0.3-0.9)

Table A 19 – Comparison of AIC values when one (participant Trio number) or two (Trio and Batch) random intercepts are used to predict cytokine concentration from pregnancy outcome using ANOVA.

Cytokine	Trio AIC	Trio and batch AIC	ANOVA P	Include batch as a level?
IL1β	145.75	113.96	6.16E-09	Yes
IL-1Ra	335.04	337.04	0.99	No
IL2	467.76	469.71	0.83	No
IL5	381.36	361.37	2.75E-06	Yes
IL6	332.22	334.06	0.68	No
IL8	357.55	357.23	0.13	No
IL10	477.68	477.18	0.11	No
IL12	394.94	377.66	1.13E-05	Yes
IL17A	388.19	365.11	5.54E-07	Yes
CCL11	161.87	159.27	0.032	Yes
FGF-basic	340.80	298.84	3.50E-11	Yes
GM-CSF	606.40	601.53	0.0088	Yes
CCL2	578.23	579.45	0.38	No
CCL3	253.87	234.00	2.93E-06	Yes
CCL4	142.67	135.02	0.0019	Yes
TNF-α	195.27	164.45	1.01E-08	Yes
VEGF	461.69	421.60	8.70E-11	Yes

Table A 20 – Unadjusted, fixed effects linear regression models predicting effect on cytokine concentration (pg/ml) from 2 or 3 group models of pregnancy outcome.

Cytokine	Model	Term	Estimate	SE	P
IL1ra	2 group model	Intercept	4.87 (4.65-5.08)	0.11	<2E-16
		Term	-0.3 (-0.58--0.02)	0.14	0.034*
	3 group model	Intercept	4.87 (4.65-5.08)	0.11	<2E-16
		Spontaneous term	-0.35 (-0.68--0.02)	0.17	0.039*
		Non spontaneous term	-0.26 (-0.58-0.06)	0.16	0.116
IL2	2 group model	Intercept	1.24 (0.88-1.61)	0.18	5.85E-10
		Term	-0.52 (-0.98--0.07)	0.23	0.026*
	3 group model	Intercept	1.24 (0.88-1.61)	0.18	5.16E-10
		Spontaneous term	-0.45 (-0.99-0.09)	0.27	0.102
		Non spontaneous term	-0.59 (-1.12--0.06)	0.27	0.030*
IL6	2 group model	Intercept	2.6 (2.38-2.83)	0.11	<2E-16
		Term	-0.34 (-0.62--0.05)	0.14	0.022*
	3 group model	Intercept	2.6 (2.38-2.83)	0.11	<2E-16
		Spontaneous term	-0.33 (-0.67-0)	0.17	0.054
		Non spontaneous term	-0.34 (-0.67--0.01)	0.17	0.047*

IL8	2 group model	Intercept	3.54 (3.29-3.8)	0.13	<2E-16
		Term	-0.29 (-0.61-0.03)	0.16	0.075
	3 group model	Intercept	3.54 (3.29-3.8)	0.13	<2E-16
		Spontaneous term	-0.26 (-0.64-0.12)	0.19	0.180
		Non spontaneous term	-0.32 (-0.69-0.05)	0.19	0.088
IL10	2 group model	Intercept	2.28 (1.89-2.66)	0.19	<2E-16
		Term	0.08 (-0.41-0.56)	0.24	0.757
	3 group model	Intercept	2.28 (1.9-2.66)	0.19	<2E-16
		Spontaneous term	-0.07 (-0.64-0.5)	0.29	0.799
		Non spontaneous term	0.21 (-0.34-0.77)	0.28	0.453
CCL2	2 group model	Intercept	1.8 (1.26-2.33)	0.27	1.33E-09
		Term	-0.66 (-1.34-0.01)	0.34	0.055
	3 group model	Intercept	1.8 (1.26-2.33)	0.27	1.22E-09
		Spontaneous term	-0.58 (-1.38-0.22)	0.40	0.154
		Non spontaneous term	-0.74 (-1.51-0.04)	0.39	0.065

Table A 21 - Unadjusted linear regression models with batch as random intercept predicting effect on cytokine concentration (pg/ml) from 2 or 3 group models of pregnancy outcome.

Cytokine			Estimate	SE	P
Il1b	2 group model	Intercept	1.17 (0.97-1.37)	0.09	1.43E-11
		Term	-0.06 (-0.21-0.09)	0.07	0.42
	3 group model	Intercept	1.17 (0.97-1.38)	0.09	6.37E-12
		Spont term	-0.11 (-0.28-0.06)	0.09	0.21
		Not spont term	-0.02 (-0.19-0.14)	0.08	0.79
Il5	2 group model	Intercept	1.41 (0.98-1.86)	0.20	4.13E-05
		Term	0.19 (-0.16-0.54)	0.18	0.31
	3 group model	Intercept	1.42 (0.98-1.87)	0.20	6.54E-05
		Spont term	0 (-0.41-0.41)	0.20	0.99
		Not spont term	0.32 (-0.06-0.7)	0.19	0.13
Il12	2 group model	Intercept	3.08 (2.64-3.52)	0.20	7.14E-12
		Term	-0.07 (-0.45-0.31)	0.19	0.72
	3 group model	Intercept	3.09 (2.65-3.51)	0.20	1.52E-12
		Spont term	-0.22 (-0.67-0.22)	0.22	0.33
		Not spont term	0.04 (-0.38-0.46)	0.21	0.84
Il17A	2 group model	Intercept	4.08 (3.62-4.56)	0.21	6.96E-10
		Term	0.13 (-0.23-0.49)	0.18	0.47
	3 group model	Intercept	4.08 (3.62-4.56)	0.21	1.27E-09
		Spont term	0.12 (-0.3-0.54)	0.21	0.59
		Not spont term	0.15 (-0.24-0.54)	0.20	0.47
CCL11	2 group model	Intercept	4.01 (3.89-4.13)	0.06	<2E-16
		Term	-0.32 (-0.46--0.16)	0.07	0.00016
	3 group model	Intercept	4.01 (3.89-4.13)	0.06	<2E-16
		Spont term	-0.35 (-0.52--0.16)	0.09	0.00027
		Not spont term	-0.29 (-0.46--0.12)	0.08	0.0014

FGF basic	2 group model	Intercept	4.15 (3.69-4.62)	0.21	3.77E-09
		Term	-0.05 (-0.34-0.23)	0.14	0.72
	3 group model	Intercept	4.15 (3.69-4.62)	0.21	3.14E-09
		Spont term	-0.05 (-0.38-0.29)	0.17	0.78
		Not spont term	-0.06 (-0.36-0.25)	0.15	0.72
GM-CSF	2 group model	Intercept	2 (1-3.07)	0.47	2.14E-04
		Term	-0.96 (-1.84--0.05)	0.42	0.032
	3 group model	Intercept	2.01 (0.99-3.09)	0.47	9.11E-05
		Spont term	-1.15 (-2.19--0.1)	0.50	0.024
		Not spont term	-0.83 (-1.78-0.14)	0.46	0.077
CCL3	2 group model	Intercept	1.72 (1.44-2.01)	0.13	5.08E-10
		Term	-0.05 (-0.28-0.19)	0.11	0.69
	3 group model	Intercept	1.72 (1.44-2.01)	0.13	2.93E-10
		Spont term	-0.07 (-0.34-0.21)	0.13	0.63
		Not spont term	-0.03 (-0.29-0.22)	0.13	0.79
CCL4	2 group model	Intercept	4.77 (4.61-4.94)	0.08	<2E-16
		Term	-0.18 (-0.34--0.02)	0.08	0.029*
	3 group model	Intercept	4.77 (4.61-4.94)	0.08	<2E-16
		Spont term	-0.13 (-0.31-0.06)	0.09	0.18
		Not spont term	-0.22 (-0.4--0.05)	0.09	0.015*
TNF-α	2 group model	Intercept	3.74 (3.51-3.97)	0.11	<2E-16
		Term	0 (-0.18-0.18)	0.09	0.99
	3 group model	Intercept	3.74 (3.51-3.97)	0.11	<2E-16
		Spont term	-0.05 (-0.26-0.15)	0.10	0.61
		Not spont term	0.04 (-0.15-0.24)	0.10	0.67
VEGF	2 group model	Intercept	2.64 (1.95-3.35)	0.32	2.89E-07
		Term	-0.19 (-0.64-0.25)	0.22	0.39
	3 group model	Intercept	2.64 (1.95-3.36)	0.32	1.56E-07
		Spont term	-0.27 (-0.79-0.24)	0.26	0.30
		Not spont term	-0.14 (-0.62-0.35)	0.24	0.58

Table A 22 – Change in IL-1Ra concentration (pg/ml) by pregnancy outcome (2 group model) adjusted for confounders, before and after removal of influential samples using a linear regression model.

	All samples			After removal of influential values*		
IL-1Ra Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	4.64 (3.77-5.51)	0.47	<2E-16	5.38 (4.79-5.98)	0.33	<2E-16
Pregnancy outcome (sPTB)						
Term	**-0.73 (-1.09--0.37)**	**0.19**	**0.00024**	**-0.38 (-0.63--0.13)**	**0.13**	**0.0054**
Batch (1)						
2	-0.1 (-0.66-0.43)	0.28	0.72	-0.12 (-0.51-0.26)	0.20	0.56
3	-0.1 (-0.64-0.44)	0.29	0.73	0.12 (-0.24-0.5)	0.20	0.54
4	-0.19 (-0.69-0.32)	0.27	0.50	-0.32 (-0.67-0.03)	0.19	0.095
5	0.57 (0.1-1.03)	0.25	0.021	0.36 (0.03-0.68)	0.17	0.041
6	0.33 (-0.13-0.76)	0.23	0.16	0.28 (-0.03-0.59)	0.17	0.092
Days before delivery at sampling	0.01 (-0.01-0.03)	0.01	0.22	0.01 (0-0.02)	0.01	0.26
Maternal BMI (normal)						
Obese	0 (-0.43-0.43)	0.23	0.99	-0.04 (-0.33-0.24)	0.16	0.79
Underweight	-0.19 (-0.91-0.53)	0.39	0.63	-0.38 (-0.86-0.1)	0.26	0.15
Smoker (No)						
Yes	-0.38 (-1.06-0.31)	0.37	0.31	-0.41 (-0.87-0.04)	0.25	0.10
Maternal ethnicity (white, white British)						
Asian, Asian British	-0.18 (-0.72-0.37)	0.29	0.55	-0.28 (-0.64-0.09)	0.20	0.16
Black, black British	-0.01 (-0.48-0.47)	0.26	0.98	-0.07 (-0.39-0.25)	0.17	0.68
All others	-0.15 (-0.5-0.2)	0.19	0.42	-0.13 (-0.36-0.1)	0.13	0.30
Maternal age at booking	0.01 (-0.02-0.03)	0.01	0.65	-0.02 (-0.04-0.01)	0.01	0.05
Maternal parity (0)						
1	0.15 (-0.18-0.48)	0.18	0.41	0.01 (-0.21-0.23)	0.12	0.91
>1	-0.03 (-0.46-0.4)	0.23	0.90	-0.01 (-0.3-0.28)	0.16	0.96

*Participants 336, 2388, 1226

Table A 23 - Change in IL-1Ra concentration (pg/ml) by pregnancy outcome (3 group model) adjusted for confounders, before and after removal of influential samples using a linear regression model.

	All samples			After removal of influential values*		
IL-1Ra Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	4.66 (3.79-5.54)	0.47	<2E-16	5.39 (4.8-5.99)	0.33	<2E-16
Pregnancy outcome (sPTB)						
Term spontaneous	**-0.8 (-1.2--0.39)**	**0.21**	**0.00033**	**-0.41 (-0.69--0.14)**	**0.15**	**0.0066**
Term non-spontaneous	**-0.66 (-1.07--0.26)**	**0.22**	**0.0027**	**-0.35 (-0.62--0.07)**	**0.15**	**0.022**
Batch (1)						
2	-0.09 (-0.65-0.43)	0.28	0.75	-0.11 (-0.5-0.26)	0.20	0.58
3	-0.09 (-0.62-0.46)	0.29	0.77	0.13 (-0.24-0.5)	0.20	0.51
4	-0.17 (-0.67-0.33)	0.27	0.53	-0.32 (-0.67-0.04)	0.19	0.10
5	0.59 (0.12-1.05)	0.25	0.018	0.37 (0.04-0.69)	0.17	0.037
6	0.35 (-0.11-0.79)	0.24	0.14	0.3 (-0.02-0.61)	0.17	0.081
Days before delivery at sampling	0.01 (-0.01-0.03)	0.01	0.23	0.01 (0-0.02)	0.01	0.27
Maternal BMI (normal)						
Obese	-0.02 (-0.46-0.41)	0.24	0.92	-0.05 (-0.34-0.24)	0.16	0.74
Underweight	-0.21 (-0.92-0.51)	0.39	0.60	-0.39 (-0.87-0.1)	0.26	0.15
Smoker (No)						
Yes	-0.36 (-1.05-0.32)	0.37	0.34	-0.41 (-0.86-0.05)	0.25	0.10
Maternal ethnicity (white, white British)						
Asian, Asian British	-0.2 (-0.75-0.35)	0.30	0.50	-0.29 (-0.66-0.08)	0.20	0.15
Black, black British	0.02 (-0.46-0.5)	0.26	0.95	-0.06 (-0.38-0.26)	0.17	0.74
All others	-0.16 (-0.51-0.19)	0.19	0.41	-0.13 (-0.37-0.1)	0.13	0.30
Maternal age at booking	0.01 (-0.02-0.03)	0.01	0.72	-0.02 (-0.04-0)	0.01	0.050
Maternal parity (0)						
1	0.14 (-0.19-0.47)	0.18	0.44	0.01 (-0.21-0.23)	0.12	0.94
>1	-0.01 (-0.44-0.43)	0.24	0.97	0 (-0.29-0.29)	0.16	0.98

*Participants 336, 2388, 1226

Table A 24 - Change in CCL2 concentration (pg/ml) by pregnancy outcome (2 group model) adjusted for confounders using a linear regression model.

	All samples			After removal of influential values		
CCL2 Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	2.43 (0.5-4.37)	1.05	0.022	-	-	-
Pregnancy outcome (sPTB)						
Term	**-1.66 (-2.47--0.86)**	**0.43**	**0.00019**	**-**	**-**	**-**
Batch (1)						
2	-1.27 (-2.51--0.05)	0.66	0.057	-	-	-
3	0.16 (-1.05-1.38)	0.66	0.80	-	-	-
4	-1.33 (-2.48--0.19)	0.62	0.034	-	-	-
5	0.79 (-0.25-1.85)	0.56	0.16	-	-	-
6	-0.26 (-1.26-0.75)	0.54	0.63	-	-	-
Days before delivery at sampling	0.01 (-0.03-0.04)	0.02	0.70	-	-	-
Maternal BMI (normal)						
Obese	-0.15 (-1.1-0.79)	0.51	0.76	-	-	-
Underweight	-2.19 (-3.79--0.6)	0.86	0.013	-	-	-
Smoker (No)						
Yes	-0.61 (-2.12-0.9)	0.82	0.46	-	-	-
Maternal ethnicity (white, white British)						
Asian, Asian British	-0.69 (-1.9-0.52)	0.66	0.29	-	-	-
Black, black British	1.15 (0.14-2.17)	0.55	0.039	-	-	-
All others	-0.51 (-1.28-0.26)	0.42	0.23	-	-	-
Maternal age at booking	0.01 (-0.05-0.06)	0.03	0.84	-	-	-
Maternal parity (0)						
1	-0.18 (-0.9-0.54)	0.39	0.65	-	-	-
>1	0.34 (-0.61-1.3)	0.52	0.51	-	-	-

Table A 25 - Change in CCL2 concentration (pg/ml) by pregnancy outcome (3 group model) adjusted for confounders using a linear regression model.

	All samples			After removal of influential values		
CLL2 Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	2.42 (0.49-4.36)	1.05	0.024	-	-	-
Pregnancy outcome (sPTB)						
Term spontaneous	**-1.63 (-2.53--0.74)**	**0.48**	**0.0010**	**-**	**-**	**-**
Term non-spontaneous	**-1.69 (-2.59--0.8)**	**0.48**	**0.0006**	**-**	**-**	**-**
Batch (1)						
2	-1.27 (-2.51--0.06)	0.66	0.057	-	-	-
3	0.16 (-1.06-1.38)	0.66	0.82	-	-	-
4	-1.34 (-2.49--0.19)	0.62	0.034	-	-	-
5	0.78 (-0.27-1.84)	0.57	0.17	-	-	-
6	-0.27 (-1.29-0.75)	0.55	0.62	-	-	-
Days before delivery at sampling	0.01 (-0.03-0.04)	0.02	0.70	-	-	-
Maternal BMI (normal)						
Obese	-0.15 (-1.1-0.81)	0.52	0.78	-	-	-
Underweight	-2.19 (-3.78--0.59)	0.87	0.013	-	-	-
Smoker (No)						
Yes	-0.62 (-2.13-0.89)	0.82	0.46	-	-	-
Maternal ethnicity (white, white British)						
Asian, Asian British	-0.68 (-1.9-0.54)	0.66	0.31	-	-	-
Black, black British	1.14 (0.12-2.16)	0.56	0.042	-	-	-
All others	-0.51 (-1.28-0.26)	0.42	0.23	-	-	-
Maternal age at booking	0.01 (-0.05-0.06)	0.03	0.83	-	-	-
Maternal parity (0)						
1	-0.17 (-0.9-0.55)	0.40	0.66	-	-	-
>1	0.33 (-0.63-1.29)	0.53	0.53	-	-	-

Table A 26 - Change in IL5 concentration (pg/ml) by pregnancy outcome (2 group model) adjusted for confounders, before and after removal of influential samples.

	All samples			After removal of influential values*		
IL5 Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	1.42 (0.47-2.37)	0.48	0.0035	1.25 (0.36-2.14)	0.45	0.0061
Pregnancy outcome (sPTB)						
Term	**0.27 (-0.1-0.64)**	**0.19**	**0.15**	**0.43 (0.07-0.78)**	**0.18**	**0.019**
Days before delivery at sampling	-0.01 (-0.03-0.01)	0.01	0.39	-0.01 (-0.03-0.01)	0.01	0.19
Maternal BMI (normal)						
Obese	-0.07 (-0.52-0.38)	0.23	0.75	-0.01 (-0.44-0.42)	0.22	0.96
Underweight	-0.37 (-1.13-0.39)	0.39	0.33	0.02 (-0.79-0.84)	0.41	0.95
Smoker (No)						
Yes	0.04 (-0.67-0.76)	0.36	0.90	0.15 (-0.54-0.83)	0.34	0.67
Maternal ethnicity (white, white British)						
Asian, Asian British	0.17 (-0.41-0.75)	0.29	0.56	0.18 (-0.36-0.73)	0.28	0.51
Black, black British	0.2 (-0.29-0.68)	0.24	0.42	0.16 (-0.33-0.65)	0.25	0.53
All others	0.1 (-0.27-0.47)	0.19	0.59	0.12 (-0.23-0.47)	0.18	0.51
Maternal age at booking	0 (-0.03-0.03)	0.01	0.93	-0.023 (-0.02-0.03)	0.17	0.87
Maternal parity (0)						
1	-0.1 (-0.45-0.24)	0.17	0.56	-0.13 (-0.46-0.2)	0.01	0.44
>1	0.34 (-0.11-0.8)	0.23	0.14	0.29 (-0.15-0.73)	0.22	0.20

*Participants 813, 1785

Table A 27 - Change in IL5 concentration (pg/ml) by pregnancy outcome (2 group model) adjusted for confounders, before and after removal of influential samples using a linear regression model.

	All samples			After removal of influential values*		
IL5 Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	1.58 (0.64-2.51)	0.47	0.0010	1.4 (0.51-2.28)	0.45	0.0022
Pregnancy outcome (sPTB)						
Term spontaneous	**0.01 (-0.4-0.41)**	**0.2**	**0.969**	**0.15 (-0.24-0.54)**	**0.20**	**0.44**
Term non-spontaneous	**0.52 (0.11-0.92)**	**0.2**	**0.012**	**0.61 (0.23-1)**	**0.20**	**0.0021**
Days before delivery at sampling	-0.01 (-0.02-0.01)	0.01	0.34	-0.01 (-0.03-0.01)	0.01	0.19
Maternal BMI (normal)						
Obese	-0.15 (-0.59-0.29)	0.22	0.51	-0.08 (-0.5-0.34)	0.21	0.70
Underweight	-0.43 (-1.17-0.31)	0.37	0.25	-0.39 (-1.1-0.32)	0.36	0.28
Smoker (No)						
Yes	0.09 (-0.61-0.78)	0.35	0.81	0.13 (-0.54-0.79)	0.33	0.70
Maternal ethnicity (white, white British)						
Asian, Asian British	-0.01 (-0.03-0.02)	0.01	0.67	0.06 (-0.48-0.59)	0.27	0.84
Black, black British	0.08 (-0.49-0.64)	0.29	0.79	0.08 (-0.38-0.54)	0.23	0.73
All others	0.26 (-0.21-0.74)	0.24	0.27	0.1 (-0.24-0.45)	0.17	0.55
Maternal age at booking	0.09 (-0.26-0.45)	0.18	0.62	-0.0016 (-0.03-0.02)	0.01	0.90
Maternal parity (0)						
1	-0.14 (-0.48-0.19)	0.17	0.40	-0.12 (-0.44-0.2)	0.16	0.46
>1	0.44 (-0.01-0.88)	0.23	0.056	0.45 (0.02-0.88)	0.22	0.038

*Participant 1785

Table A 28 - Change in CCL11 concentration (pg/ml) by pregnancy outcome (2 group model) adjusted for confounders, before and after removal of influential samples using a linear regression model.

	All samples			After removal of influential values*		
CCL11 Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	3.83 (3.41-4.25)	0.21	<2E-16	3.85 (3.42-4.27)	0.21	<2E-16
Pregnancy outcome (sPTB)						
Term	**-0.32 (-0.47--0.16)**	**0.08**	**0.000056**	**-0.24 (-0.41--0.07)**	**0.084**	**0.0049**
Days before delivery at sampling	0.0018 (-0.01-0.01)	0	0.67	0 (-0.01-0.01)	0.0042	0.97
Maternal BMI (normal)						
Obese	0.03 (-0.19-0.24)	0.11	0.81	0.04 (-0.18-0.25)	0.11	0.73
Underweight	0.03 (-0.33-0.4)	0.18	0.85	-0.01 (-0.36-0.34)	0.18	0.96
Smoker (No)						
Yes	-0.04 (-0.39-0.3)	0.17	0.80	-0.03 (-0.38-0.31)	0.17	0.85
Maternal ethnicity (white, white British)						
Asian, Asian British	-0.15 (-0.42-0.12)	0.14	0.28	-0.15 (-0.42-0.12)	0.14	0.26
Black, black British	0 (-0.24-0.23)	0.12	0.97	0.03 (-0.2-0.26)	0.12	0.79
All others	-0.12 (-0.3-0.05)	0.09	0.17	-0.14 (-0.32-0.03)	0.089	0.11
Maternal age at booking	0.01 (-0.01-0.02)	0.01	0.38	0.0043 (-0.01-0.02)	0.0064	0.50
Maternal parity (0)						
1	-0.03 (-0.2-0.14)	0.08	0.73	-0.02 (-0.18-0.15)	0.083	0.82
>1	0.03 (-0.19-0.25)	0.11	0.79	0.04 (-0.17-0.25)	0.11	0.71

*Participant 1375

Table A 29 - Change in CCL11 concentration (pg/ml) by pregnancy outcome (3 group model) adjusted for confounders, before and after removal of influential samples using a linear regression model.

	All samples			After removal of influential values*		
CCL11 Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	3.84 (3.42-4.26)	0.21	<2E-16	3.85 (3.42-4.28)	0.22	<2E-16
Pregnancy outcome (sPTB)						
Term spontaneous	**-0.34 (-0.51--0.15)**	**0.09**	**0.00015**	**-0.24 (-0.44--0.05)**	**0.094**	**0.011**
Term non-spontaneous	**-0.3 (-0.48--0.12)**	**0.09**	**0.0013**	**-0.24 (-0.43--0.05)**	**0.10**	**0.014**
Days before delivery at sampling	0 (-0.01-0.01)	0.00	0.68			
Maternal BMI (normal)				0 (-0.01-0.01)	0.0042	0.97
Obese	0.02 (-0.2-0.24)	0.11	0.85	0.04 (-0.18-0.25)	0.11	0.74
Underweight	0.03 (-0.33-0.39)	0.18	0.86	-0.01 (-0.36-0.34)	0.18	0.95
Smoker (No)						
Yes	-0.04 (-0.38-0.3)	0.17	0.81	-0.03 (-0.38-0.31)	0.18	0.85
Maternal ethnicity (white, white British)						
Asian, Asian British	0.01 (-0.01-0.02)	0.01	0.40	-0.15 (-0.42-0.12)	0.14	0.26
Black, black British	-0.16 (-0.43-0.12)	0.14	0.26	0.03 (-0.2-0.26)	0.12	0.79
All others	0 (-0.23-0.23)	0.12	1.00	-0.14 (-0.32-0.03)	0.089	0.11
Maternal age at booking	-0.12 (-0.3-0.05)	0.09	0.17	0 (-0.01-0.02)	0.0064	0.50
Maternal parity (0)						
1	-0.03 (-0.2-0.13)	0.08	0.70	-0.02 (-0.18-0.15)	0.083	0.82
>1	0.04 (-0.19-0.26)	0.11	0.75	0.04 (-0.18-0.26)	0.11	0.71

* Participant 1375

Table A 30 - Change in GM-CSF concentration (pg/ml) by pregnancy outcome (2 group model) adjusted for confounders using a linear regression model.

	All samples			After removal of influential values		
GM-CSF Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	2.98 (0.69-5.28)	1.16	0.011	-	-	-
Pregnancy outcome (sPTB)						
Term	**-1.06 (-2.01--0.09)**	**0.45**	**0.020**	**-**	**-**	**-**
Days before delivery at sampling	0.02 (-0.03-0.06)	0.02	0.46	-	-	-
Maternal BMI (normal)						
Obese	0.32 (-0.81-1.45)	0.57	0.57	-	-	-
Underweight	-2.01 (-3.86--0.16)	0.94	0.034	-	-	-
Smoker (No)						
Yes	1.18 (-0.55-2.92)	0.88	0.18	-	-	-
Maternal ethnicity (white, white British)						
Asian, Asian British	0.02 (-1.38-1.42)	0.71	0.98	-	-	-
Black, black British	0.09 (-1.11-1.3)	0.60	0.88	-	-	-
All others	-0.24 (-1.15-0.67)	0.46	0.60	-	-	-
Maternal age at booking	-0.03 (-0.1-0.03)	0.03	0.32	-	-	-
Maternal parity (0)						
1	-0.17 (-1.01-0.67)	0.42	0.69	-	-	-
>1	0.87 (-0.31-2.04)	0.59	0.15	-	-	-

Table A 31 - Change in GM-CSF concentration (pg/ml) by pregnancy outcome (3 group model) adjusted for confounders using a linear regression model.

	All samples			After removal of influential values		
GM-CSF Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	3.08 (0.78-5.39)	1.17	0.0092	-	-	-
Pregnancy outcome (sPTB)						
Term spontaneous	**-1.27 (-2.33--0.19)**	**0.50**	**0.013**	**-**	**-**	**-**
Term non-spontaneous	**-0.87 (-1.91-0.21)**	**0.51**	**0.091**	**-**	**-**	**-**
Days before delivery at sampling	0.02 (-0.03-0.06)	0.02	0.48	-	-	-
Maternal BMI (normal)						
Obese	0.24 (-0.9-1.38)	0.57	0.67	-	-	-
Underweight	-2.06 (-3.9--0.21)	0.94	0.030	-	-	-
Smoker (No)						
Yes	1.21 (-0.51-2.95)	0.87	0.17	-	-	-
Maternal ethnicity (white, white British)						
Asian, Asian British	-0.04 (-0.1-0.03)	0.03	0.28	-	-	-
Black, black British	-0.05 (-1.46-1.36)	0.71	0.94	-	-	-
All others	0.14 (-1.06-1.36)	0.60	0.81	-	-	-
Maternal age at booking	-0.23 (-1.14-0.68)	0.46	0.61	-	-	-
Maternal parity (0)						
1	-0.21 (-1.04-0.63)	0.42	0.63	-	-	-
>1	0.95 (-0.24-2.13)	0.60	0.11	-	-	-

Table A 32 - Change in CCL4 concentration (pg/ml) by pregnancy outcome (2 group model) adjusted for confounders using a linear regression model

	All samples			After removal of influential values		
CCL4 - Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	4.87 (4.44-5.3)	0.22	<2E-16	-	-	-
Pregnancy outcome (sPTB)						
Term	**-0.18 (-0.35--0.01)**	**0.08**	**0.034**	**-**	**-**	**-**
Days before delivery at sampling	0 (-0.01-0.01)	0	0.82	-	-	-
Maternal BMI (normal)						
Obese	-0.01 (-0.23-0.2)	0.11	0.92	-	-	-
Underweight	-0.25 (-0.6-0.1)	0.18	0.17	-	-	-
Smoker (No)						
Yes	-0.04 (-0.38-0.31)	0.17	0.83	-	-	-
Maternal ethnicity (white, white British)						
Asian, Asian British	-0.15 (-0.42-0.12)	0.14	0.27	-	-	-
Black, black British	0.2 (-0.03-0.43)	0.12	0.083	-	-	-
All others	0.17 (0-0.35)	0.09	0.052	-	-	-
Maternal age at booking	0 (-0.02-0.01)	0.01	0.69	-	-	-
Maternal parity (0)						
1	-0.08 (-0.24-0.08)	0.08	0.34	-	-	-
>1	-0.08 (-0.3-0.13)	0.11	0.43	-	-	-

Table A 33 - Change in CCL4 concentration (pg/ml) by pregnancy outcome (3 group model) adjusted for confounders using a linear regression model.

	All samples			After removal of influential values*		
CCL4 - Predictor (baseline)	β (95% CI)	β SE	P	β (95% CI)	β SE	P
Intercept	4.84 (4.41-5.27)	0.22	<2E-16	4.83 (4.41-5.26)	0.21	<2E-16
Pregnancy outcome (sPTB)						
Term spontaneous	**-0.13 (-0.32-0.06)**	**0.09**	**0.16**	**-0.1 (-0.29-0.09)**	**0.093**	**0.30**
Term non-spontaneous	**-0.23 (-0.42--0.04)**	**0.10**	**0.016**	**-0.22 (-0.41--0.03)**	**0.095**	**0.022**
Days before delivery at sampling	0.00081 (-0.01-0.01)	0.00	0.85	0 (-0.01-0.01)	0.0041	0.78
Maternal BMI (normal)						
Obese	0.0033 (-0.21-0.22)	0.11	0.98	0 (-0.21-0.22)	0.11	0.97
Underweight	-0.24 (-0.59-0.12)	0.18	0.19	-0.04 (-0.43-0.36)	0.20	0.85
Smoker (No)						
Yes	-0.05 (-0.39-0.3)	0.17	0.79	-0.02 (-0.36-0.32)	0.17	0.91
Maternal ethnicity (white, white British)						
Asian, Asian British	-0.13 (-0.4-0.14)	0.14	0.34	0.26 (0.03-0.5)	0.12	0.030
Black, black British	0.19 (-0.04-0.42)	0.12	0.11	0.18 (0.01-0.35)	0.087	0.042
All others	0.18 (0-0.35)	0.09	0.048	0 (-0.01-0.01)	0.0063	0.79
Maternal age at booking	-0.0016 (-0.01-0.01)	0.01	0.80	-0.11 (-0.37-0.16)	0.13	0.42
Maternal parity (0)						
1	-0.07 (-0.23-0.09)	0.08	0.40	-0.1 (-0.26-0.07)	0.082	0.24
>1	-0.1 (-0.32-0.11)	0.11	0.35	-0.15 (-0.37-0.07)	0.11	0.17

* Participant 813

Academic Publications

Placenta 46 (2016) 31–37

Contents lists available at ScienceDirect

Placenta

journal homepage: www.elsevier.com/locate/placenta

A new biological and clinical resource for research into pregnancy complications: The Baby Bio Bank

Lydia J. Leon [a, *], Nita Solanky [a], Susanne E. Stalman [b], Charalambos Demetriou [a], Sayeda Abu-Amero [a], Philip Stanier [a], Lesley Regan [c], Gudrun E. Moore [a]

[a] UCL Institute of Child Health, 30 Guilford Street, London WC1N 1EH, United Kingdom
[b] Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands
[c] Obstetrics and Gynaecology Department, St Mary's Hospital, Imperial College, Praed Street, London W2 1NY, United Kingdom

ARTICLE INFO

Article history:
Received 15 March 2016
Received in revised form
27 July 2016
Accepted 22 August 2016

Keywords:
Biobank
Pregnancy complications
Preterm birth
Fetal growth restriction
Miscarriage
Pre-eclampsia

ABSTRACT

About 20% of pregnancies are affected by some form of complication. Research has shown that anomalies in implantation, development, and growth of the fetus; ineffective nutrient exchange between mother and fetus due to placental dysfunction; and maternal problems such as hypertension or infection during pregnancy can all lead to adverse pregnancy outcomes. However, the molecular aetiology of such events remains poorly understood. Fetal growth restriction (FGR), recurrent miscarriage (RM), preterm birth (PTB), and pre-eclampsia (PE) are the most common pregnancy complications encountered in the UK and these outcomes can result in an array of morbidities in both mother and baby, and in the most severe cases in mortality. We need to know more about normal pregnancy and where the important triggers are for failure. This prompted us to collect a large set of biological samples with matching clinical data from over 2500 normal and abnormal pregnancies, for use in research into these conditions. This paper outlines the nature of these sample sets and their availability to academia and industry, with the intention that their widespread use in research will make significant contributions to the improvement of maternal and fetal health worldwide (http://www.ucl.ac.uk/tapb/sample-and-data-collections-at-ucl/biobanks-ucl/baby-biobank).

1. Introduction

Each year, tens of thousands of women in the UK, and millions across the world, encounter mild to life-threatening pregnancy complications. Despite substantial developments in the clinical management and biological understanding of such cases, many questions regarding the molecular aetiology of common complications in pregnancy remain unanswered. A key challenge often associated with research into pregnancy is obtaining sufficient numbers of biological samples to conduct well-powered studies. The Baby Bio Bank (BBB) is a large, UK based biobank that was set up to remove the obstacle of recruitment and to support research into the environmental and genetic mechanisms underlying common complications of pregnancy. The BBB provides ethically approved, high-quality biological samples and clinical data to academic and industrial partners, both nationally and internationally. Funded primarily through the UK based charity, *Wellbeing of Women*, the BBB began recruiting pregnant mothers from three hospitals across London in 2009. The project's recruitment phase has now reached completion, amassing over 54,000 biological samples from 2515 healthy and complicated pregnancies.

Recruitment and sample collection was based at three London hospitals: Queen Charlotte and Chelsea (QC), Chelsea and Westminster (CW), and St Mary's (SM). Together these London maternity services deliver around 13,000 infants a year. The BBB received ethical approval for collection from these hospitals from Trent Derby Research Ethics Committee [1]. Ethical consent for biological sample collection from the proband (fetus/infant), mother, and father, as well as access to participating patient records was obtained in advance of the birth. All samples are stored at the UCL Institute of Child Health, with a duplicate of the entire biobank located at SM, Imperial College London. Semi-anonymised patient data, downloaded directly from electronic clinical records and supplemented by data gathered by the BBB recruitment team (see **S6** for questionnaire format used), is stored in a secure, custom

* Corresponding author.
E-mail address: lydia.leon.11@ucl.ac.uk (L.J. Leon).

http://dx.doi.org/10.1016/j.placenta.2016.08.085

designed online database that links directly to matching sample information.

Fetal growth restriction (FGR), pre-eclampsia (PE), preterm birth (PTB) and recurrent miscarriage (RM) were chosen as the primary foci for the BBB on account of their prevalence within the UK, but also because they capture a physiological spectrum of pregnancy disorders. These disorders are all considered multifactorial, caused by the combined and varying effects of the maternal and fetal genotype, the intrauterine environment, and many other clinical and environmental factors. It is thought likely that many issues covering abnormal implantation to preterm membrane rupture and delivery may share similar or overlapping aetiologies, including physiological mechanisms, pathways and even specific genetic predisposition variants.

The BBB contains data and paired biological samples from 236 FGR, 133 PE, 373 PTB, and 232 RM pregnancies. Over 1500 'normal' pregnancies with none of the above complications were also collected for use as control samples, and 636 of these are classified as 'perfect' controls with no recorded problems associated with the mothers' health, pregnancy, or delivery.

A unique and valuable aspect of this dataset is the comprehensiveness of the paternal samples and clinical data that has been amassed, accompanying 68% of the participating pregnancies. Availability of these samples enables investigators interested in genetic contributions to common pregnancy complications to conduct powerful genetic studies using traditional trio designs, or to study in isolation the paternal contribution to pregnancy and fetal outcome.

This paper summarises the design, recruitment, and main demographic and clinical characteristics of the BBB cohort, and provides researchers with contact details to enquire about availability for sample groups of interest. Detailed information regarding sample collection, preparation, and storage are available in Supplementary Information.

2. Recruitment design

On account of the expertise of the BBB's founders and the likely high contribution of genetic factors to the phenotypes of interest, the BBB was principally conceptualised as a genetic and epigenetic resource. However, additional opportunities for biomarker research are available due to the parallel collection of plasma/sera from mothers, fathers and infants, in addition to urine from the mother.

Recruitment for the BBB was carried out over a four-year period at three hospitals (SM, CW, QC), following a targeted prospective cohort design (Fig. 1). Where possible, recruitment occurred at the antenatal clinic or at follow-up appointments when conditions of interest first presented. This normally gave participants at least several days to ask questions about the consent form and information leaflets. Some women were also recruited on antenatal and labour wards. Any consenting family having a baby at one of the participating units was able to contribute.

Given the ambitious size of the project, a variety of strategies were employed to ensure that recruitment to each complication was as high as possible, whilst still ensuring that they were recruited prior to delivery. Many women were targeted for recruitment if deemed to be at risk for a particular complication. For example, women belonging to certain risk groups (e.g. with a hypertensive disorder, or complicated pregnancy history) were approached by recruiters. Nurses and clinicians would also notify recruiters when women began presenting with clinical symptoms of a complication, regardless of prior risk status (e.g. at the onset of pre-eclampsia or following preterm membrane rupture). In addition, a large number of normal, uncomplicated pregnancies were also recruited for comparative purposes, with some of these pregnancies ultimately contributing to the complications groups (e.g. a woman with no known complication at the time of recruitment could end up delivering preterm).

This process was designed to be feasible within a routine clinical setting. Whereas a simple randomised or stratified randomised design would have been theoretically preferable in minimising selection bias, this was not possible given the context and the resources at hand.

The BBB recruitment team consisted of three full-time recruiters: two research associates with international experience in recruitment and research, and one research nurse with substantial experience of recruitment to large cohort collections, all of whom were trained phlebotomists and had expertise in the necessary sample preparation techniques. Each recruiter was based at one of the three hospitals taking part in the project. The recruiters were responsible for counselling and obtaining consent from the parents, as well as collecting and preparing samples around the clock. Recruitment numbers for each hospital are listed in Table 1. Recruitment was initiated at SM, before recruitment commenced at CW, followed by QC.

3. Sample collection and storage

The variety of samples (and relevant accompanying information) that were intended for collection from each pregnancy are presented in Table 2. Recruiters focused on securing the collection of 'Trio' sample sets, in which tissue, DNA, and RNA are available for mother, father, and baby. This aim was achieved for 1328 pregnancies in total, across all of the complications and control pregnancies. If we assume a dominant model of inheritance for a potential genetic trait of interest, and a significance threshold of 5%, these trio numbers within the BBB would have reasonable power (>0.7) to detect risk variants with relative risks above 1.5, if all cases are combined into one larger 'complications group' (under the hypothesis that certain variants may affect all complications), or 2–2.5 if case groups are considered individually (see **S5** for further details).

Each recruitment centre took responsibility for sample receipt and storage. Samples received were matched with the clinical phenotype data and barcoded at point of entry. All tissue and blood samples are stored at −80 °C until requested, with tissue stored in *RNAlater*. DNA and RNA are stored at −80 °C and are available for all samples in which the relevant primary tissue is available.

Each placenta was collected at birth and processed as quickly as possible (normally within one hour). Four 1 cm^3 sections were excised from beneath the placental membrane, equidistant from the umbilical cord and washed in PBS to remove excess maternal blood. Collecting multiple samples from each placenta increased the total amount of tissue available, as well as facilitating investigations into variation in gene/protein expression in different parts of placenta. By taking samples from four distinct sites, mosaicism can be detected by standard karyotyping. Placental mosaicism is present in approximately 2% of chorionic villous samples [2] and it is likely that a number of such samples will be present in our collection. In addition to sampling from the chorionic plate, villous tissue from the maternal side of the placenta, as well as umbilical cord, cord blood, and fetal membrane tissue were also collected where possible. All tissue was immediately transferred to a vile containing *RNAlater* to minimise degradation of nucleic acids. Occasionally the placenta, cord or cord blood were not available (e.g. a baby was delivered at home, in an emergency, or another hospital). In such instances a buccal swab from the baby was requested and DNA extracted from this specimen. Buccal swabs for DNA extraction were also taken from fathers where blood was unavailable.

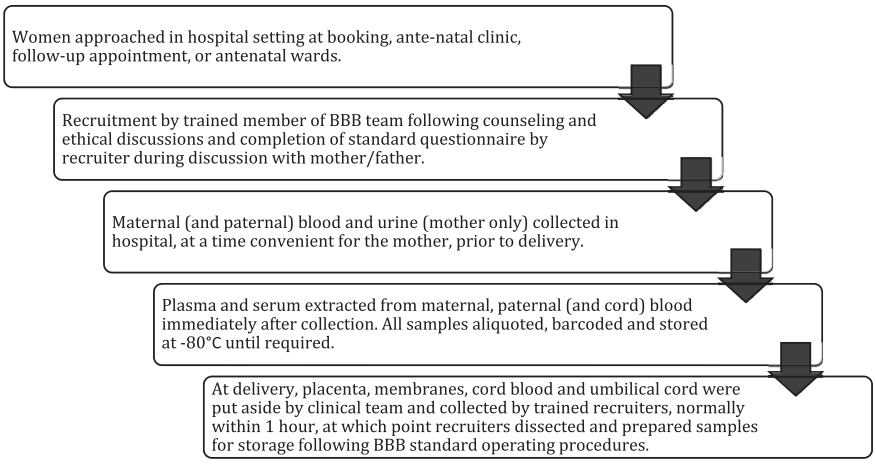


Fig. 1. Flow-diagram outlining BBB recruitment and sample collection process.

Table 1
Hospital recruitment rates.

Hospital	N (%)
St Mary's	1107 (44)
Chelsea and Westminster	880 (35)
Queens Charlotte's and Chelsea	528 (21)

Table 2
Variety of samples collected from participating trios.

Sample type
Maternal whole blood/serum/plasma
Paternal whole blood/serum/plasma
Maternal urine
Maternal DNA and RNA (from blood/buccal swab)
Paternal DNA and RNA (from blood/buccal swab)
Placental parenchyma, villous, and membrane
Cord blood
Umbilical cord tissue
Baby DNA and RNA (from placenta/buccal swab)

For ease of collection and to encourage participation by minimising the number of blood draws/hospital visits for mothers, maternal blood and urine samples were collected once during pregnancy, at a time that was convenient to the mother, and were not restricted to a certain time point in gestation. This usually coincided with hospital appointments and blood draws being taken for clinical purposes. Plasma and serum were extracted immediately from such samples (see **S2** for detailed protocol). Where possible, approximately 20 ml of blood was collected from each of the parents. Whole blood, serum and plasma were aliquoted into maximum volumes of 2 ml to minimise exposure of samples to freeze-thaw cycles. Urine was collected in standard specimen containers by the donor and aliquoted into 2 ml volumes within 1 h, for long-term storage at −80 °C. Given the expected high yields of DNA from the protocols used for extraction, and the volume of samples collected, tens of thousands of standard molecular analyses will be possible using the amount of DNA and RNA available within the bank.

Maternal blood collection ranged from 7 weeks gestation until delivery, with the mean gestational age for collection being 25 weeks. Bloods are available for points collected throughout pregnancy, with peaks at certain times, such as week 12 which coincides with a key antenatal hospital visit for mothers in the UK (Fig. 2). 20% of blood samples were collected within the 1st trimester, 38% in the 2nd, and 42% in the 3rd. BBB sample requests can be specified by gestational age. As maternal samples were collected once from each pregnancy, consecutive samples are not available. We recognise that this, and the variation in time points of blood collection, would be a limitation for certain biomarker studies, however this maximised the number of participants within the study.

To ensure biological samples were processed and maintained to the highest possible standards necessary for use across all common molecular applications, detailed quality control audits using PCR, sequencing, and nucleic acid integrity assessment with the Agilent 2200 TapeStation System, have been conducted on a random selection of at least 15% of the whole cohort. These tests have assured high purity and integrity of DNA and RNA extracted from BBB samples, supporting their use in standard molecular biology assays including Sanger sequencing, qPCR and end-point PCR in which all samples tested positive for housekeeping genes used as standards. A number of samples have also been used successfully for exome analyses and genome-wide methylation assays, in which high DNA quality and integrity is essential.

Both informal and formal audits of the BBB were conducted. Informal audits were conducted by the BBB manager. Formal audits were carried out annually by both the UCL and Imperial Human Tissue Authority committees respectively. On all audits there was 100% concordance with samples in the electronic database and the physical location of samples. This was seen in both directions (from database to sample and sample to database). Retrieval of consent forms for all samples was also 100%.

4. Clinical and demographic characteristics

Pregnancy related morbidity and mortality are known to be associated with various neonatal and maternal characteristics that are well documented in the BBB, and may be of interest either as the central phenotype under investigation or in downstream multivariable analyses. To ensure the BBB clinical database was as comprehensive as possible, data were collected by recruiters using standardised forms in addition to detailed clinical downloads from electronic maternity records at each hospital. The two collections were subsequently merged by the BBB data manager into a format that is available to BBB users. This design enabled the collection of certain additional maternal data (e.g. occupation) as well as paternal data, which were not available on clinical records. This strategy also allowed for discrepancies in data entry to be identified and corrected via comparison of any duplicate categories. A

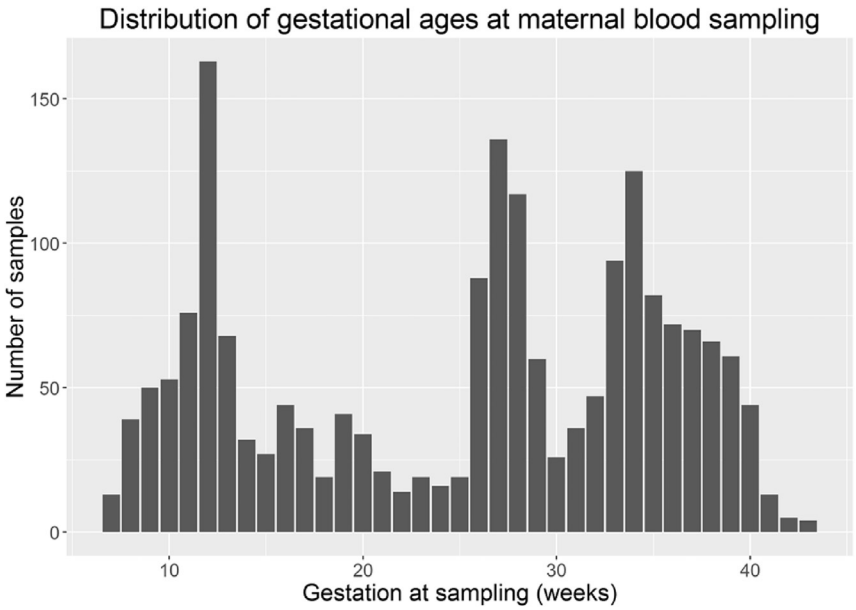


Fig. 2. Graph showing distribution of gestational ages at maternal blood sampling within the BBB.

selection of the key recorded information and their provenances are listed in Table 3. A full list of clinical data categories is available on request.

The original project aimed to collect only pregnancies from white European mothers and as can be seen (Table 4) this remains the largest group within the BBB. However, due to the multinational population in London it proved difficult to identify and consent a strictly white European cohort. Many clinics were attended by individuals of Asian and African origin expressing an interest in participating in the BBB. We were aware that such a valuable resource would be of greater value if it were inclusive of all ethnic backgrounds and ultimately included such individuals in the hope that funding would one day be available to cover a similarly sized cohort collection for these ethnic groups too.

Ethnicity can be grouped in numerous ways and there is often a lack of consensus within clinical and research communities on precisely how to do this. The BBB electronic database contains two separate data columns on maternal ethnicity, that are used widely throughout the NHS and are based on those used in the national census. Ethnicity data is summarized into four broad categories to give the reader an idea of the ethnic distribution within the BBB cohort. The main demographic characteristics of currently available clinical data of the cohort are outlined in Table 4. Some categories, e.g. delivery method, have larger amounts of missing data than others. The further population of these missing fields is ongoing, requiring the interrogation of paper rather than electronic notes.

Table 3
Selection of clinical information available.

Maternal	Paternal	Fetal
Age[a,b]	Age[a]	Gender[b]
BMI[b]	Height[a,c]	Gestation[b]
Height [b,c]	Weight[a,c]	Birthweight[b]
Weight [b,c]	BMI[c]	Placental Weight[c]
Ethnicity[b]	Ethnicity[a]	Head Circumference[b]
Pre-Pregnancy Weight[a]	Diabetes[a]	Birth Length[b]
Parity[a,b]	Hypertension[a]	Apgar Score[b]
Diabetes[a,b]	Smoking[a]	Mode Of Delivery[b]
Hypertension[a,b]	Medication[a]	Congenital abnormalities[b]
Smoking[a,b]	Occupation[a]	
Medication[a,b]		
Occupation[a,b]		
Pregnancy history[a,b]		
Age at menarche[a,b]		
Marital status[a,b]		
Infectious disease[b]		

[a] Information volunteered from patient to recruitment staff.
[b] Information gathered from clinical records.
[c] Measured by recruiting staff.

5. Definitions and verification of outcome data.

Clinical diagnosis of FGR, PE, RM and PTB is recorded for all our samples, the definitions for each of these complications were consistent across the three NHS hospitals from which our samples were collected and are outlined below. However, we recognise that researchers may have specific definitions to categorise clinical complications. To address this, each BBB sample is linked to, and can be searched by more than 200 fields of data that we hold, enabling bespoke categorisation of sample sets, as well as a deeper analysis of data by the creation of sub-categories as required. Whilst this potential for sub-categorisation is a valuable feature of the Bank, we recognise that any such manipulation of the cohort is limited by the original clinical definitions used during recruitment, as stipulated in the following sections.

6. Preterm birth

Prematurity affects between 5 and 18% of births worldwide and is the leading cause of neonatal death globally [3]. In the UK, around 7% of births are preterm [4]. Prematurity is associated with a complex array of morbidities from neurodevelopmental issues, to gastrointestinal complications that often extend beyond the neonatal period, through childhood and into adulthood [5]. Preterm births in the BBB were defined as any delivery occurring before 37 weeks gestation.

Biological samples were collected from the 373 preterm pregnancies within the BBB, 122 of which were trios. To achieve this sizeable collection, recruiters actively targeted mothers with a history of preterm birth, women presenting with spontaneous

Table 4
Demographic characteristics of the BBB cohort.

Variable	Category	N (%)	Range (SD)	Mean
Maternal ethnicity	White	1565 (68.2)		
	Black	229 (10.0)		
	Asian	99 (4.3)		
	Other[b]	401 (17.5)		
Maternal smoking	Non-smoker	2074 (87.7)		
	Smoker	100 (4.2)		
	Quit within last 12 months	190 (8.1)		
Maternal age at booking	<20	25 (1.3)	14–55 (5.48)	32.83
	20–25	187 (9.6)		
	26–30	372 (19.1)		
	31–35	741 (38.0)		
	36–40	478 (24.5)		
	41–45	140 (7.2)		
	>46	6 (0.3)		
Maternal BMI at booking	<18.5 (Underweight)	71 (3.0)	14–66 (5.27)	24.87
	18.5–24.9 (Normal weight)	1304 (55.3)		
	25.0–29.9 (Pre-obesity)	630 (26.7)		
	30.0–34.9 (Obesity Class 1)	217 (9.2)		
	35–39.9 (Obesity Class 2)	94 (4.0)		
	Above 40 (Obesity Class 3)	42 (1.8)		
Maternal parity	Nulliparous	1324 (53.5)	0–12 (0.95)	0.67
	Primiparous	797 (32.2)		
	Multiparous	354 (14.3)		
Neonate gender	Male	1182 (52.0)		
	Female	1092 (48.0)		
Number of infants	Singletons	2445 (97.2)		
	Twins	70 (2.7)		
Delivery method	CS	716 (42.2)		
	Vaginal	980 (57.8)		
Birthweight (g)			295–5470 (728.92)	3164
GA at birth (weeks)[a]			20–43 (2.81)	38.32

[a] Assessed from ultra-sound scanning.
[b] Chinese, other Asian, other black, other, and all mixed groups.

preterm membrane rupture or threatening preterm labour, women with shortened cervices, or those with a positive fetal fibronectin result. Research into the underlying aetiology of preterm birth often focuses on the most extreme 'very early preterm' cases in order to best elucidate underlying mechanisms. Generally classified as those born at less than 33 weeks gestation, these births are also the hardest to collect biological material from, given the complicated nature of such deliveries. The BBB has managed to secure biological and clinical data from 96 such valuable cases.

The mean gestational age at birth for this group (Table 5) is 33 weeks and the youngest delivery was at just 20 weeks gestation, 3 weeks below the current 'limit of viability' upheld by obstetricians [6]. PTB is often split into three main phenotypic categories for research: spontaneous, indicated and preterm premature rupture of membranes (PPROM). The numbers of pregnancies within each of these sub-categories are also displayed in Table 5. This sub-categorisation was carried out with reference to data from clinical records that provided labour, membrane rupture and delivery information. Definitions used for categorisation are listed below Table 5. The BBB preterm deliveries were fairly evenly distributed between the three main modes of delivery: emergency caesarean section (CS), elective CS, and vaginal.

7. Pre-eclampsia

PE is a multisystem disorder that affects around 2–8% of pregnancies globally [7–9]. Patients typically present with hypertension during pregnancy in conjunction with proteinuria. The condition tends to affect women in the latter half of pregnancy and is associated with both FGR and PTB. PE is estimated to account for 10–15% of all global maternal mortality, with women in developing countries bearing the vast majority of this burden [10].

In our cohort, pre-eclampsia was diagnosed as new hypertension, i.e. at least two consecutive blood pressure (BP) readings above 140/90 mmHg or an increase in systolic BP of at least 30 mmHg or diastolic BP of at least 15 mmHg above booking, combined with new proteinuria defined as a protein creatinine ratio (PCR) of greater than 50, or 24 h quantitation with a level of greater than 300 mg. These diagnoses were recorded in the clinical notes and available to BBB recruiters and data managers for subsequent categorisation. Women who had personal or family history of PE, displayed one or more of the symptoms of the disorder, or were being actively monitored or treated for it were targeted by the recruitment team.

The BBB recruited 133 women who went on to have a pre-eclamptic pregnancy, from which over 700 biological samples were accrued. 41 of these pregnancies were collected as trios. 45 of these women also delivered preterm, representing an important sub-phenotype of the syndrome. As well as women with clinically diagnosed PE, the BBB also contains a substantial number of women with PE associated symptoms such as pregnancy induced hypertension (PIH) or renal complications (as identified from maternal clinical records). The number of pregnancies associated with these morbidities within the BBB are summarized in Table 6.

8. Fetal growth restriction

FGR describes the condition of a fetus unable to reach its growth potential [11,12], which often results in a small for gestational age (SGA) baby. SGA is defined as a birthweight for gestational age below the 10th centile and is often used as a proxy for FGR during pregnancy. In high-income countries, FGR affects about 5–6% of pregnancies [13–16]. FGR is associated with perinatal morbidity and mortality. Furthermore, survivors are at risk for adult-onset

Table 5
Number of preterm pregnancies according to gestational age and mode of delivery.

Preterm birth data (cumulative N)									
Gestational age at birth (weeks)	Total	Spontaneous PTB[a]	Spontaneous onset of labour[b]	PPROM[c]	Indicated PTB	Missing data on labour/ membrane	Caesarean section delivery[d]	Vaginal delivery	Induced labour
20	1	1	0	1	0	0	0	0	1
21	1	0	0	1	0	1	0	0	1
22	2	0	0	1	0	2	0	0	1
23	4	0	3	4	0	4	0	3	1
24	6	0	4	5	1	5	0	5	2
25	16	8	6	8	2	6	2	7	2
26	15	11	9	10	4	0	6	8	2
27	30	17	16	15	6	7	9	12	3
28	33	22	19	20	11	0	17	14	3
29	41	24	21	21	17	0	24	15	3
30	61	30	27	25	21	10	31	17	3
31	80	37	33	31	32	11	40	20	3
32	96	44	39	36	40	12	49	25	3
33	132	58	51	47	60	14	73	32	3
34	186	80	70	64	86	20	94	43	6
35	252	102	87	85	124	26	121	68	21
36	373	145	113	123	186	42	172	105	48

[a] Spontaneous labour and/or membrane rupture and delivery before 37 weeks.
[b] Labour is spontaneous and delivery occurs before 37 weeks.
[c] Membrane rupture is spontaneous and resultant delivery occurs before 37 weeks.
[d] Emergency and elective.

Table 6
Number of pregnancies associated with a pre-eclamptic phenotype.

Pre-eclampsia phenotype	BBB N
Pre-eclampsia	133
PIH	57
PIH in previous pregnancy/ies	76
Essential hypertension	98
Cardiac complications	15
Renal complications	16
Thrombosis	8

diseases, such as type 2 diabetes, obesity and cardiovascular disease [17–20].

Although the original intention was to establish a pure FGR cohort, we have chosen to use a broader definition of growth restriction in this summary paper, and report here the characteristics of SGA infants within the BBB, whose birthweight for age centile (as defined by the 1996 gender specific four-in-one growth charts produced by the Child Growth Foundation, London) fell below the 10th centile. Whilst we are aware of the limitations that this definition holds as a proxy for all FGR pregnancies, with the current debate surrounding definitions of FGR and variations in clinical diagnoses we decided to use this looser definition under the assumption that a large number of these SGA babies would indeed be true cases of FGR. Furthermore, more detailed clinical data from the pregnancy, including from ultra-sound scans that enable the mapping of growth trajectories, can be retrieved if a more conservative and strict definition of FGR is required by BBB users.

The BBB recruited 234 women with babies that fell below the 10th birthweight for gestational age centile. 82 of these deliveries were successfully recruited as trios. Within this group, 112 babies were below the 5th birthweight centile and 50% delivered preterm. The mean birthweight for this SGA cohort was 1880g. The distribution of birthweight for age centiles within the BBB, alongside the mean birthweight for each centile grouping are summarized in Table 7.

Alongside interest in the physiological, genetic and environmental underpinnings of FGR, there is now a growing interest in deliveries at the other end of the spectrum in which neonates are large for gestational age, known as macrosomia. Over 200 babies in the BBB were born above the 95th birthweight for age centile and over 300 were above the 90th centile. This group poses particular risks to mothers during delivery and is an important phenotype to study in attempts to minimise maternal morbidity and mortality related to childbirth.

9. Recurrent miscarriage

Miscarriage is the commonest complication of pregnancy and is defined as the spontaneous loss of a fetus before it has reached viability. Hence, the term miscarriage includes all the losses from conception up to the 24th week of gestation. Recurrent miscarriage affects about 1.5% of couples trying to conceive [21]. Studies have shown that the risk of miscarriage increases after each successive pregnancy loss and it can reach 45% after three consecutive losses [22]. Recurrent miscarriage (RM) was defined as three or more consecutive pregnancy losses, data that was collected from the clinical records.

The BBB contains samples and clinical data for 232 women who have a clinical diagnosis of RM, as well as many who have experienced one or two previous miscarriages (Table 8). A large number of these cases were recruited from the internationally recognized centre for mothers' suffering from recurrent miscarriage based at SM and run in part by the BBB principal investigator, LR. 97 of the RM pregnancies belong to a BBB trio. In total over 1500 miscarriages were experienced by women in the cohort, and 770

Table 7
Distribution of centiles and mean birthweight across the BBB.

Centile	BBB N	Mean birthweight (g)
<5th	112	1555
5th-<10th	122	2179
10th -< 25th	576	2614
25th-<50th	410	2876
50th-<75th	401	3273
75th-<90th	273	3590
90th-<95th	143	3913
>95th	206	4122
Data not available	272	NA

Table 8
Distribution of number of previous miscarriages within BBB.

No. previous miscarriages	BBB N
0	1666
1	399
2	139
3	121
4	55
5	32
6	11
7	7
8	2
9	3
10	1
Data not available	79

Table 9
Summary of extra clinical categories of interest.

Maternal clinical category	BBB N
Diabetes	23
Gestational Diabetes	31
Pregnancy from In vitro Fertilization	22
Placenta previa	11
Placental abruption	26
Urinary tract infection	36
Thyroid complications	27
Cardiac complications	29
Epilepsy	48

women had had at least one previous miscarriage prior to their BBB pregnancy. 20 women were receiving treatment with aspirin during their pregnancy.

10. Other complications

Alongside the four main complications in pregnancy, the size of the BBB means that a substantial number of other maternal complications were also recorded, as detailed in Table 9.

11. Using the BBB

Interested parties are encouraged to contact the BBB management team with any questions or requests they may have relating to the bank. Procurement of samples is subject to review by the BBB Research Management Board (who have so far approved all 14 initial applications), as well as written evidence from the applicant of an ethically approved project for which the samples will be used. Enquiries should be directed to the BBB Manager, Dr Nita Solanky (nita.solanky@ucl.ac.uk). Further information on the BBB including our SOPs, contact details and application form are available at http://www.ucl.ac.uk/tapb/sample-and-data-collections-at-ucl/biobanks-ucl/baby-biobank.

Conflicts of interest

None.

Acknowledgements

The BBB team would like to thank Wellbeing of Women as well as all participating hospital staff and families. This research was funded by the UK Medical Research Council, Biological Research Council and Great Ormond Street Hospital Children's Charity (LJL PhD fellowship), and supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children, NHS Foundation Trust, and University College London.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.placenta.2016.08.085.

References

[1] S. Abu-Amero, A. Thomas, S. White, K. Rogers, A. Miranda, N. Solanky, L.J. Leon, C. Demetriou, X. Ke, S. Stanier, B. Stanier, H. Costello, S. Tzehaie, L. Al-Olabi, C. Williamson, M. Johnson, L. Regan, G.E. Moore, The Baby Bio Bank-A legacy for researchers worldwide into common complications of pregnancy, J. General Pract. 2 (3) (2014).
[2] D.K. Kalousek, M. Vekemans, Confined placental mosaicism, J. Med. Genet. 33 (7) (1996) 529–533.
[3] H. Blencowe, S. Cousens, D. Chou, M. Oestergaard, L. Say, A.B. Moller, M. Kinney, J. Lawn, the Born Too Soon Preterm Birth Action G., Born Too Soon: the global epidemiology of 15 million preterm births, Reprod. health 10 (Suppl 1) (2013) S2.
[4] Office for National Statistics, Gestation-Specific Infant Mortality, 2012, p. 2014.
[5] S. Saigal, L.W. Doyle, An overview of mortality and sequelae of preterm birth from infancy to adulthood, Lancet 371 (9608) (2008) 261–269.
[6] I. Seri, J. Evans, Limits of viability: definition of the gray zone, J. Perinatol. 28 (S1) (2008) S4–S8.
[7] WHO, Geographic variation in the incidence of hypertension in pregnancy. World health organization international collaborative study of hypertensive disorders of pregnancy, Am. J. Obstet. Gynecol. 158 (1) (1988) 80–83.
[8] C.L. Roberts, J.B. Ford, C.S. Algert, S. Antonsen, J. Chalmers, S. Cnattingius, M. Gokhale, M. Kotelchuck, K.K. Melve, A. Langridge, C. Morris, J.M. Morris, N. Nassar, J.E. Norman, J. Norrie, H.T. Sørensen, R. Walker, C.J. Weir, Population-based trends in pregnancy hypertension and pre-eclampsia: an international comparative study, BMJ Open 1 (1) (2011).
[9] WHO, WHO Recommendations for Prevention and Treatment of Pre-eclampsia and Eclampsia, World Health Organization, Geneva, 2011.
[10] L. Duley, The global impact of pre-eclampsia and eclampsia, Semin. Perinatol. 33 (3) (2009) 130–137.
[11] ACOG, Intrauterine growth restriction. Clinical management guidelines for obstetrician-gynecologists, Int. J. Gynaecol. Obstet. 72 (1) (2001) 85–96.
[12] S.L. Hillman, S. Finer, M.C. Smart, C. Mathews, R. Lowe, V.K. Rakyan, G.A. Hitman, D.J. Williams, Novel DNA methylation profiles associated with key gene regulation and transcription pathways in blood and placenta of growth-restricted neonates, Epigenetics 10 (1) (2015) 50–61.
[13] D. Brodsky, H. Christou, Current concepts in intrauterine growth restriction, J. intensive care Med. 19 (6) (2004) 307–319.
[14] C. Demetriou, S. Abu-Amero, A.C. Thomas, M. Ishida, R. Aggarwal, L. Al-Olabi, L.J. Leon, J.L. Stafford, A. Syngelaki, D. Peebles, K.H. Nicolaides, L. Regan, P. Stanier, G.E. Moore, Paternally expressed, imprinted insulin-like growth factor-2 in chorionic villi correlates significantly with birth weight, PloS One 9 (1) (2014) e85454.
[15] G.E. Moore, M. Ishida, C. Demetriou, L. Al-Olabi, L.J. Leon, A.C. Thomas, S. Abu-Amero, J.M. Frost, J.L. Stafford, Y. Chaoqun, A.J. Duncan, R. Baigel, M. Brimioulle, I. Iglesias-Platas, S. Apostolidou, R. Aggarwal, J.C. Whittaker, A. Syngelaki, K.H. Nicolaides, L. Regan, D. Monk, P. Stanier, The role and interaction of imprinted genes in human fetal growth, Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 370 (1663) (2015) 20140074.
[16] R. Harding, A.D. Bocking, Fetal Growth and Development, Cambridge University Press, Cambridge, 2001.
[17] T. Forsen, J. Eriksson, J. Tuomilehto, A. Reunanen, C. Osmond, D. Barker, The fetal and childhood growth of persons who develop type 2 diabetes, Ann. Intern. Med. 133 (3) (2000) 176–182.
[18] R.A. Simmons, Developmental origins of adult disease, Pediatr. Clin. N. Am. 56 (3) (2009) 449–466.
[19] D.J. Barker, Fetal growth and adult disease, Br. J. Obstet. Gynaecol. 99 (4) (1992) 275–276.
[20] D.J. Barker, C.N. Hales, C.H. Fall, C. Osmond, K. Phipps, P.M. Clark, Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth, Diabetologia 36 (1) (1993) 62–67.
[21] G.M. Stirrat, Recurrent miscarriage, Lancet 336 (8716) (1990) 673–675.
[22] L. Regan, P.R. Braude, P.L. Trembath, Influence of past reproductive performance on risk of spontaneous abortion, BMJ 299 (6698) (1989) 541–545.

PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Review

Cite this article: Moore GE *et al*. 2015 The role and interaction of imprinted genes in human fetal growth. *Phil. Trans. R. Soc. B* **370**: 20140074.
http://dx.doi.org/10.1098/rstb.2014.0074

One contribution of 13 to a discussion meeting issue 'Human evolution: brain, birthweight and the immune system'.

Subject Areas:
genetics, health and disease and epidemiology, molecular biology

Keywords:
genomic imprinting, fetal growth restriction, placenta, chorionic villus sampling, birth weight, type 1 diabetes

Author for correspondence:
Gudrun E. Moore
e-mail: gudrun.moore@ucl.ac.uk

†Present address: Plant and Microbial Biology Department, University of California Berkeley, Berkeley, CA 94720-3102, USA.
‡Present address: Neonatal Unit, Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona 08908, Spain.
§Present address: UCL EGA Institute for Women's Health, London WC1E 6AU, UK.
¶Present address: Imprinting and Cancer Group, Epigenetics and Cancer Biology Program, Bellvitge Institute for Biomedical Research, L'Hospitalet de Llobregat, Barcelona 08908, Spain.

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2014.0074 or via http://rstb.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

The role and interaction of imprinted genes in human fetal growth

Gudrun E. Moore[1], Miho Ishida[1], Charalambos Demetriou[1], Lara Al-Olabi[1], Lydia J. Leon[1], Anna C. Thomas[1], Sayeda Abu-Amero[1], Jennifer M. Frost[1,†], Jaime L. Stafford[1], Yao Chaoqun[1], Andrew J. Duncan[1], Rachel Baigel[1], Marina Brimioulle[1], Isabel Iglesias-Platas[1,‡], Sophia Apostolidou[1,§], Reena Aggarwal[1], John C. Whittaker[2], Argyro Syngelaki[3], Kypros H. Nicolaides[3], Lesley Regan[4], David Monk[1,¶] and Philip Stanier[1]

[1]Genetics and Epigenetics in Health and Diseases Section, Genetics and Genomic Medicine Programme, UCL Institute of Child Health, London WC1N 1EH, UK
[2]Noncommunicable Disease Epidemiology Unit, London School of Hygiene and Tropical Medicine, University of London, London WC1E 7HT, UK
[3]Harris Birthright Research Centre for Fetal Medicine, King's College Hospital, London SE5 9RS, UK
[4]Department of Obstetrics and Gynaecology, Imperial College London, St Mary's Campus, London W2 1NY, UK

Identifying the genetic input for fetal growth will help to understand common, serious complications of pregnancy such as fetal growth restriction. Genomic imprinting is an epigenetic process that silences one parental allele, resulting in monoallelic expression. Imprinted genes are important in mammalian fetal growth and development. Evidence has emerged showing that genes that are paternally expressed promote fetal growth, whereas maternally expressed genes suppress growth. We have assessed whether the expression levels of key imprinted genes correlate with fetal growth parameters during pregnancy, either early in gestation, using chorionic villus samples (CVS), or in term placenta. We have found that the expression of paternally expressing insulin-like growth factor 2 (*IGF2*), its receptor *IGF2R*, and the *IGF2*/*IGF1R* ratio in CVS tissues significantly correlate with crown–rump length and birthweight, whereas term placenta expression shows no correlation. For the maternally expressing pleckstrin homology-like domain family A, member 2 (*PHLDA2*), there is no correlation early in pregnancy in CVS but a highly significant negative relationship in term placenta. Analysis of the control of imprinted expression of *PHLDA2* gave rise to a maternally and compounded grandmaternally controlled genetic effect with a birthweight increase of 93/155 g, respectively, when one copy of the *PHLDA2* promoter variant is inherited. Expression of the growth factor receptor-bound protein 10 (*GRB10*) in term placenta is significantly negatively correlated with head circumference. Analysis of the paternally expressing delta-like 1 homologue (*DLK1*) shows that the paternal transmission of type 1 diabetes protective G allele of rs941576 single nucleotide polymorphism (SNP) results in significantly reduced birth weight (−132 g). In conclusion, we have found that the expression of key imprinted genes show a strong correlation with fetal growth and that for both genetic and genomics data analyses, it is important not to overlook parent-of-origin effects.

1. Background and results

(a) Fetal growth

Birthweight and its relationship to mortality show one of the strongest links observed in epidemiology, illustrated by a reverse-J-shaped curve with the highest mortality observed in the lightest and heaviest groups [1]. Growing appropriately *in utero* is essential for a long and healthy life. Fetal growth

restriction (FGR) affects approximately 6% of pregnancies, and is identified in approximately half of stillborn fetuses without malformations [2,3]. While the majority of FGR babies demonstrate catch-up growth, the combination of suboptimal intrauterine growth followed by accelerated childhood growth can increase their susceptibility to adult-onset diseases, including type 2 diabetes, hypertension and coronary artery disease [4]. Each baby's unique growth potential *in utero* is determined by the successful nutritional and respiratory support from the mother to the fetus via a placenta, and disturbing this balance could lead to FGR [5]. Fetal growth is influenced by both genetic and environmental factors, although the relevant molecular pathways are still poorly defined. Identifying key genes and pathways that regulate fetal growth will allow for better monitoring of intrauterine growth, maximizing healthy outcomes.

(b) Genomic imprinting

Genomic imprinting is a process of epigenetic modification on the genome that causes silencing of one allele according to its parental origin, resulting in monoallelic expression, without changing the DNA sequence [6–8]. Sex-specific imprint marks are heritable to daughter cells, but are erased and re-established in the germline during gametogenesis [9]. Evidence from mouse models and rare human imprinting disorders suggests that genes that are paternally expressed tend to increase fetal growth, whereas maternally expressed genes restrict fetal growth. For example, mice knockouts for paternally expressed genes *Igf2*, mesoderm-specific transcript (*Mest*) and paternally expressed gene 3 (*Peg3*) result in FGR, whereas mice deficient for maternally expressed genes insulin-like growth factor 2 receptor (*Igf2r*), *H19* and *Grb10* show an overgrowth phenotype [10–14] (table 1). Rare imprinting disorders such as the growth-restricted phenotype of Silver–Russell syndrome (SRS) may implicate complex roles involving both absence of growth promoters such as *IGF2* and potential increase of growth restrictors such as *GBR10* (reviewed in [26]).

The kinship theory or parental 'conflict theory' predicts that imprinting may have evolved as a result of competition between the paternal and maternal genome for maternal nutrient provision. The paternal genome encourages fetal growth by extracting nutrients from the mother, whereas the maternal genome counterbalances this by limiting resources to the offspring to ensure not only her survival, but also the equal provision of nutrients among her offspring [41]. Genomic imprinting is observed predominantly in placental mammals, and it is, indeed, the placenta which serves as the key regulatory site for this genomic conflict.

More than 100 imprinted genes have been identified in mice and approximately half of them are conserved in humans. In addition to this, many more tissue-specific human-imprinted loci are being discovered (http://igc.otago.ac.nz/; http://www.har.mrc.ac.uk/) [42]. In the current project, we have studied 13 imprinted genes that are highly expressed in human term placenta and are known to lead to growth phenotypes when deficient in mice (table 1). In addition, we included three non-imprinted genes that were critical to the action of *IGF2*, which is a key paternally expressed imprinted growth promoter (table 2). We have investigated the expression of these genes in both early and late gestation using the King's College London (KCL) CVS cohort (11–13 weeks of gestation) and the Moore term placenta cohort, respectively, and correlated these data with important growth parameters such as birthweight, placental weight and head circumference.

Also, in a separate analysis reported here, the potential influence of other variables such as the baby's sex, gestational age, parity, maternal weight/body mass index (BMI) and maternal smoking were tested against gene expression. In some situations, loss of imprinting (LOI) can occur, leading to biallelic expression of the gene. Because this could potentially influence the overall gene dosage, term placenta and CVS samples used in these expression studies were also investigated to see whether they retained a normal imprinting pattern, or showed monoallelic expression. In this hybrid review/research article, we summarize our previous findings together with new data.

(c) Insulin-like growth factor axis and *IGF2/H19* locus

The insulin/IGF growth factor 'axis' constitutes key regulatory endocrine factors of pre- and postnatal growth. These include insulin (*INS*), *IGF1*, *IGF2* and their corresponding receptors (*IR*, *IGF1R* and *IGF2R*), and six binding proteins (*IGFBP1–6*) [48]. INS and IGF1 exclusively bind to IR and IGF1R, respectively, whereas IGF2 can bind to IGF1R, IGF2R and IR 11-isoform [49]. *IGF2R* is located on human chromosome 6q25.3 and shows maternal expression in only 10% of term placentas and CVS [17,50]. One of its major functions is the lysosomal targeting and degradation of IGF2, thus acting as a growth suppressor [51]. *IGF2* and *H19* map to one of the most intensely studied imprinted gene clusters on human chromosome 11p15. Their reciprocal imprinting is controlled by differential methylation of imprinting control region 1 (ICR1) which is normally only methylated on the paternal allele [52]. The unmethylated maternal ICR1 allows the binding of the CTCF transcription factor, blocking the access of *IGF2* promoters to the *H19* downstream enhancers, resulting in the activation of *H19* expression. Conversely, the CTCF protein is prevented from binding to the paternal methylated ICR1, resulting in monoallelic paternal *IGF2* expression owing to *IGF2* promoter interaction with the enhancers. Approximately 50% of the growth-restricted SRS cases show loss of methylation at ICR1, which could lead to decreased *IGF2* expression [24] and that may well contribute to SRS growth restriction.

In our previous studies, we have shown that *IGF2* and *IGF2R* expression in term placenta has no correlation with baby's birth size parameters. However, their expression levels in CVS tissues showed a strong positive correlation with birthweight [17,33], indicating their role as 'early growth effectors'. In addition to this study, the expression levels of *H19* ($n = 104$) relative to the ribosomal protein L19 (*L19*) endogenous control gene in CVS tissues was measured by RT-quantitative polymerase chain reaction (qPCR). The relative expression levels of *H19* were correlated to birth weight in a regression model adjusted for baby's sex, parity, gestational age at birth, maternal BMI and smoking habits. The CVS expression data for *IGF2*, *IGF2R*, *H19*, *PHLDA2*, *IGF1* and *IGF1R* were also correlated to CRL at the gestational age of 12 weeks, using the same regression model, except this time the gestational age at CRL measurement was used instead of gestational age at birth. Correlation between *H19* expression and birthweight was not statistically significant ($p = 0.07$). However, there was significant evidence for positive association between CRL at 12 weeks and *IGF2* expression ($p = 0.004$; figure 1*a*), *IGF2R* expression ($p = 0.03$; figure 1*b*),

Table 1. Imprinted genes highly expressed in the placenta. Origin, parental origin of the expressed allele; M, maternally expressed; P, paternally expressed; ncRNA, non-coding RNA; FGR, fetal growth restriction; Dup, duplication; UPD, uniparental disomy; ICR, imprinting control region; LBW, low birthweight; BW, birthweight; HC, head circumference; CVS, chorionic villus sampling tissues; CRL, crown–rump length; PIP, phosphatidylinositol phosphate lipid; mat del, maternally inherited deletion; pat del, paternally inherited deletion; T1D, type 1 diabetes; TNDM, transient neonatal diabetes mellitus; BWS, Beckwith–Wiedemann syndrome; SRS, Silver–Russell syndrome; CNV, copy number variation; asterisk, findings from this study.

locus	gene	origin	description	mouse KO phenotypes	human growth phenotypes
6q24	*PLAGL1*	P	zinc finger protein	FGR, bone malformation, high neonatal lethality [15]	TNDM (pUPD6, pDup6q24, ICR hypomethylation) [16]
6q25	*IGF2R*	M/biallelic	clearance of IGF2	fetal and placental overgrowth, organ and skeletal abnormalities [11]	CVS expression positively correlated to BW [17] and CRL*
7p12	*GRB10*	M/P	GF receptor-bound protein	fetal and placental overgrowth [10]	implicated in SRS (mDup7p11.2–13) [18]; term placenta expression negatively associates with HC*
7q21.3	*PEG10*	P	retrotransposon derived	embryonic lethal due to placental malformation [19]	hypermethylation at ICR and reduced expression in LBW cord blood [20]; upregulated in FGR placenta [21]
7q32.2	*MEST*	P/biallelic	α/β hydrolase fold family	fetal and placental growth restriction, high postnatal lethality, abnormal maternal behaviour [12]	implicated in SRS (mUPD 7q31-qter) [22]
11p15	*H19*	M	long ncRNA	fetal and placental overgrowth [13,23]	ICR1 hypomethylation [24] and CNV [25] in SRS
	IGF2	P	growth factor	fetal and placental growth restriction	CVS expression positively correlated to BW [17] and CRL*; implicated in BWS and Wilm's tumour [26]
	CDKN1C	M	tumour suppressor	gestational fetal and placental overgrowth [27]	mutated in IMAGe [28], BWS [29] and SRS [30] patients
	SLC22A18	M	organic cation transporter	not reported	term placenta expression associated with HC [31]
	PHLDA2	M	PH domain, PIP binding	placental overgrowth [32]	highly expressed in lower BW and FGR placenta [21,33–35]; promoter variant associated with BW [36]
14q32	*DLK1*	P	transmembrane glycoprotein	pre- and postnatal growth restriction, high perinatal lethality, obese postnatally [37]	associated with T1D [38], UPD14 syndromes [26], T1D SNP correlated to BW*
	MEG3	M	ncRNA	postnatal lethal (mat del), pre- and postnatal growth restriction, high perinatal lethality (pat del) [39]	associated with T1D [38], reduced expression in FGR placenta [35]
19q13.4	*PEG3*	P	zinc finger protein	placental and fetal growth restriction, abnormal maternal behaviour [14]	tumour suppressor [40]

IGF2/*IGF1R* ratio ($p = 0.03$; figure 1*c*) and *H19* expression ($p = 0.04$; figure 1*d* and table 3). These results suggest that the many members of the IGF axis (*IGF2*, *IGF2R* and *IGF1R*), and the closely associated *H19*, shape the growth trajectory early in pregnancy.

There was no correlation between maternal smoking and the expression in CVS of the genes tested (those listed above) in our samples. Nevertheless, we observed an association between *IGF2* expression and parity, whereby *IGF2* expression is higher in the 'parity greater than one' group of babies ($p = 0.03$; electronic supplementary material, figure S1*a*); this is consistent with the role of *IGF2* as a positive growth regulator. This observation is interesting as the majority of second born babies are bigger [36]. We also found evidence that the maternal BMI was positively correlated with *IGF2R* expression ($p = 0.03$; electronic supplementary material,

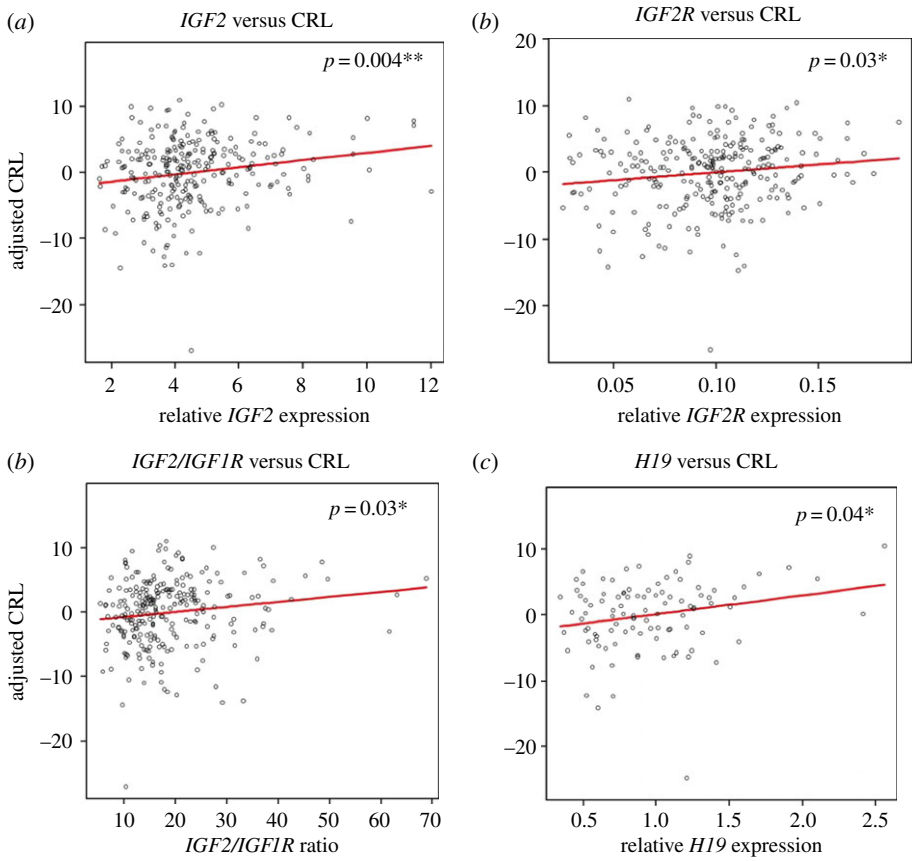


Figure 1. Correlation between imprinted gene expression in CVS and CRL. Expression levels of each gene relative to the *L19* endogenous control gene were correlated to crown–rump length (CRL: mm) using a multiple linear regression model adjusted for maternal BMI, baby's sex, parity, gestational age when CRL was measured and maternal smoking habit. Positive correlations with CRL and (*a*) *IGF2* expression ($r = 0.77$; $p = 0.004$), (*b*) *IGF2R* expression ($r = 0.76$; $p = 0.03$), (*c*) *IGF2/IGF1R* ratio ($r = 0.74$; $p = 0.03$) and (*d*) *H19* expression ($r = 0.74$; $p = 0.04$) were observed. (Online version in colour.)

Table 2. Non-imprinted genes highly expressed in the placenta.

locus	gene	description	mouse KO phenotypes	human growth phenotypes
7p12	*IGFBP3*	carrying protein for IGF1 and IGF2	retinal vessel loss [43]	implicated in common cancers [44]
12q23.2	*IGF1*	growth promoter	pre- and postnatal growth restriction, infertile [45]	pre- and postnatal growth restriction [46]
15q26.3	*IGF1R*	IGF1 and IGF2 receptor	fetal growth restriction and perinatal lethal	pre- and postnatal growth restriction [47]

figure S1*b*) and negatively correlated with *IGF1* ($p = 0.046$; electronic supplementary material, figure S1*c*). This suggests that it is important to allow for correction for maternal BMI/weight when investigating the gene expression in association with fetal growth. Interestingly, *H19* was expressed significantly higher in males ($p = 0.006$; electronic supplementary material, figure S1*e* and table S2). The observed sex bias cannot be explained by LOI (i.e. biallelic expression of *H19* in males only), because all of the CVS tissues tested retained monoallelic expression (table 4), therefore it is likely to result from upregulation of the active maternal copy. Males are normally born bigger than females [36], and the sexual dimorphism in antenatal biometry has been reported to be evident around 8–12 weeks of gestation [53]. As *H19* is a negative growth regulator, the higher expression may help prevent male babies from growing too large.

(d) *GBR10*

GRB10 is located in the human chromosome 7q12 imprinted region. Chromosome 7 is implicated in causality for SRS, because 10% of patients show maternal uniparental disomy of chromosome 7. FGR is a key feature of SRS, which has been suggested to result either from the overexpression of a maternally expressed gene or loss of a paternally expressed growth-promoting gene. *GRB10* encodes a growth factor receptor binding protein that can interact with receptor tyrosine kinases and intracellular proteins [54]. *GRB10* is imprinted in an isoform- and a tissue-specific manner [55]. In humans, *GRB10* shows biallelic expression in most tissues, while exhibiting isoform-specific paternal expression in the brain but with maternal expression confined to the placental villous trophoblast [56]. In mice, *Grb10* is paternally expressed in the brain, but shows ubiquitous maternal expression in other tissues [55]. This pattern is roughly

Table 3. The association between mRNA levels and fetal growth in term placenta and CVS. Shading indicates previously published results [17,33]. The correlation significance is indicated by *p*-values. Correlation coefficient (*r*) is presented underneath the *p*-values for the associations reaching significance. *n*, number of samples; BW, birth weight; PW, placental weight; HC, head circumference; CRL, crown–rump length; NT, not tested.

	CVS			term placenta			
gene	*n*	BW	CRL	*n*	BW	PW	HC
IGF2	260	0.009**	0.004** ($r = 0.77$)	200	0.9	0.46	0.43
IGF2R	260	0.004**	0.03* ($r = 0.76$)	200	0.86	0.56	0.7
IGF2/IGF2R	260	0.93	0.58	200	0.5	0.56	0.62
IGF1	200	0.48	0.07	NT	NT	NT	NT
IGF1R	260	0.08	0.93	NT	NT	NT	NT
IGF1/IGF1R	200	0.76	0.06	NT	NT	NT	NT
IGF2/IGF1R	260	0.005**	0.03* ($r = 0.74$)	NT	NT	NT	NT
PHLDA2	260	0.55	0.92	200	0.0001**	0.7	0.95
MEST	NT	NT	NT	200	0.96	0.78	0.42
H19	104	0.07	0.04* ($r = 0.74$)	86	0.28	0.42	0.51
DLK1	99	0.67	0.25	272	0.07	0.8	0.4
GRB10	NT	NT	NT	193	0.64	0.69	0.04* ($r = -0.35$)
MEG3	NT	NT	NT	195	0.88	0.54	0.43
PEG10	NT	NT	NT	110	0.48	0.21	0.32
PEG3	NT	NT	NT	93	0.87	0.52	0.82
SLC22A18	NT	NT	NT	78	0.13	0.57	0.52
CDKN1C	NT	NT	NT	81	0.82	0.6	0.98
PLAGL1	NT	NT	NT	102	0.77	0.91	0.61
PLAGL1imp	NT	NT	NT	102	0.18	0.55	0.064
IGFBP3	NT	NT	NT	102	0.63	0.62	0.49

Table 4. Summary of imprinting analysis in CVS tissues and term placenta. M, maternal expression; P, paternal expression. %, percentage of samples with monoallelic expression within informative samples; n.a., not available.

gene	parental origin	imprinting in term placenta	imprinting in CVS	polymorphic site
IGF2	P	67/67 (100%)	40/40 (100%)	rs680
IGF2R	M/biallelic	n.a.	3/24 (12%)	rs1805075
PHLDA2	M	11/11 (100%)	21/21 (100%)	rs13390, rs1056819
MEST	P/biallelic	34/42 (81%)	n.a.	rs10863
H19	M	19/19 (100%)	33/33 (100%)	rs2067051
DLK1	P	30/30 (100%)	n.a.	rs1802710
MEG3	M	9/9 (100%)	n.a.	rs45617834, rs941575
PEG3	P	14/16 (88%)	n.a.	rs1055359
PEG10	P	42/42 (100%)	n.a.	rs13073, rs13226637
GRB10	M (placenta), P (brain)	n.a.	n.a.	n.a.
SLC22A18	M	23/23 (100%)	n.a.	rs1048046, rs1048047
PLAGL1	P	11/11 (100%)	n.a.	rs2076684
CDKN1C	M	24/24 (100%)	n.a.	PAPn repeat

the opposite of what is seen for *Igf2*, where it is preferentially maternally expressed in the adult mouse brain but paternally expressed in other tissues [57]. Inactivation of the maternal copy of *Grb10* results in fetal and placental overgrowth, indicative of its role as a potent growth suppressor [10]. In contrast, mice with a disrupted paternal copy showed normal growth but increased social dominance behaviour, illustrated by increased facial barbering (whisker removal) on cage-mates [58].

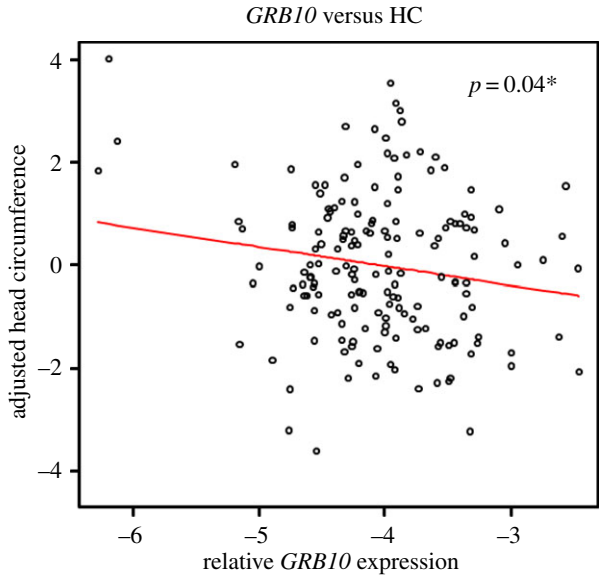


Figure 2. Negative correlation between *GRB10* term placental expression and head circumference. The expression level of *GRB10* relative to the *L19* housekeeping gene was correlated to head circumference (cm) using a multiple linear regression model adjusted for baby's sex, parity, gestational age at birth, maternal weight and smoking habits. *GRB10* expression values in logarithmic scale was used. Significant negative association was observed for *GRB10* term placenta expression and head circumference ($r = -0.35$; $p = 0.04$). (Online version in colour.)

In this study, we have observed a significant negative association between *GRB10* expression (all isoforms) and head circumference (figure 2, $p = 0.04$), but no significant correlation with birthweight ($p = 0.64$) or placental weight ($p = 0.69$; table 3). The direction of association is consistent with the role of *GRB10* as a negative growth regulator. It is interesting that the observed association is specific to head circumference, because it is oppositely imprinted in the brain. There was no correlation between maternal smoking and expression of the genes tested in term placenta samples (electronic supplementary material, table S3). Interestingly, *GRB10* expression showed a positive association with increasing gestational age ($p = 0.03$; electronic supplementary material, figure S2*a*). This suggests that *GRB10* is acting to suppress the head circumference of the baby close to birth, because a head size too large for the birth canal would be detrimental for the mother.

(e) *PHLDA2*

PHLDA2 is a maternally expressed gene located on the centromeric domain of the Chr11p15 imprinting cluster, along with other maternally expressed genes *CDKN1C* and *SLC22A18*. *PHLDA2* encodes a small (144 amino acid) protein with a Pleckstrin-homology (PH) domain which has the capacity to bind membrane phosphatidylinositol phosphate lipids (PIPs) [59], suggesting a role for it as a cell signalling protein. In line with the kinship theory, *Phlda2*-deficient mice have an enlarged placenta, whereas overexpression of *Phlda2* in transgenic mice results in placental stunting with a modest reduction in fetal weight [60,61]. We have previously shown that birth weight is not correlated with *PHLDA2* expression levels in CVS tissues, but has a significant negative correlation in term placenta [17,33], indicative of a function as a 'late growth effector'. Other studies have observed upregulation of *PHLDA2* in FGR placentas [21,34,35], and in first and second trimester miscarriage placentas [62]; these data all support the hypothesis that *PHLDA2* is an important negative regulator of growth.

More recently, upregulation of placental *PHLDA2* expression among mothers who smoke during pregnancy has been reported [63]. In our study, however, we did not observe any correlation between maternal smoking and CVS or term placental expression of *PHLDA2* (electronic supplementary material, table S3). *PHLDA2* expression in CVS and term placenta did not show correlation with any of the confounding variables used in the model, except for gestational age. We identified that reduced *PHLDA2* expression in CVS tissues was associated with advancing gestational age at birth ($p = 0.0092$; electronic supplementary material, figure S1*e*). Because a shorter gestation results in smaller babies, its high expression in CVS fits its role as a growth suppressor.

All the samples used in the analysis showed monoallelic expression of *PHLDA2*, demonstrating that LOI cannot account for the increased expression seen in the smaller birth weight babies [33]. To further investigate this correlation, we successively interrogated the nearby region for potential genetic variations that correlate with fetal growth. We identified a rare 15 bp repeat sequence variant (RS1) in the *PHLDA2* promoter region, which has been shown to reduce the *PHLDA2* promoter efficiency [36]. Maternal inheritance of RS1 resulted in a 93 g increase in birthweight, and when the mother is homozygous for RS1, the effect on birthweight is 155 g, suggesting a grand-maternal influence. Paternal inheritance of RS1 does not influence fetal growth as the variant lies on the epigenetically silenced paternal allele, emphasizing the importance of taking into account parent-of-origin effects when analysing genetic variants. Taken together, these data show that *PHLDA2* is a strong negative growth suppressor and provide a potential pre-pregnancy test, using the RS1 variant, to predict birthweight.

(f) *DLK1*

DLK1 (*PREF1* and *FA1*) is a paternally expressed gene located in the human chromosome 14q32 imprinting cluster, approximately 90 kb away from the maternally expressed non-coding RNA gene *MEG3* (also called *GTL2*). *DLK1* encodes a transmembrane glycoprotein with six epidermal growth factor-like repeat motifs [64], known to be involved in adipogenesis [65]. *Dlk1*-null mice show high perinatal lethality, pre- and postnatal growth restriction followed by an obese phenotype [37], suggesting that it acts as a growth promoter.

In this study, the expression levels of *DLK1* (all isoforms) in CVS ($n = 99$) and term placenta ($n = 272$) were correlated to fetal growth parameters. For the CVS analysis, only the tissues from extreme birthweight babies (less than 10th centile and more than 90th centile) were used. Using the regression model as described for *H19*, we did not observe any association between *DLK1* expression and birthweight ($p = 0.23$) or with CRL ($p = 0.16$). However, term placental *DLK1* expression did show a weak positive association with birthweight ($p = 0.07$; table 3). Although this trend did not reach statistical significance, the direction of influence is consistent with its role as a growth promoter. Interestingly, *DLK1* expression showed a positive correlation with increasing parity ($p = 0.05$; electronic supplementary material, figure S2*b* and table S3), possibly increasing the size of the later parity babies.

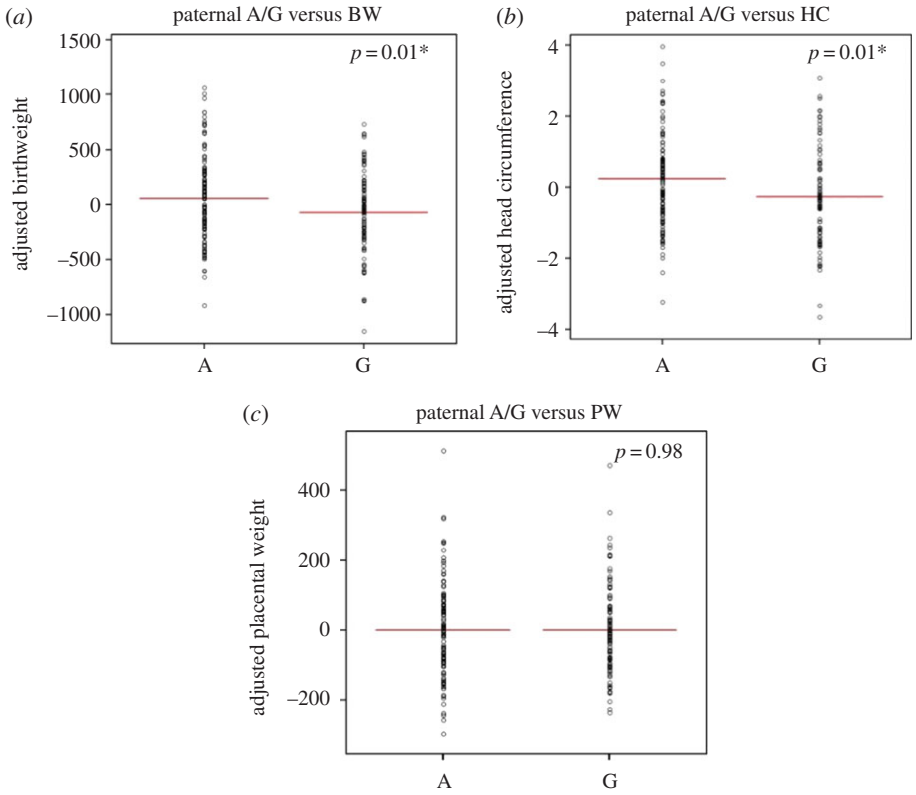


Figure 3. The association between paternal A/G SNP rs941576 at the *DLK1* locus and fetal growth. Partial residual plots illustrating the correlation between paternal inheritance of the A or G allele and (*a*) birthweight (g), (*b*) head circumference (cm) and (*c*) placental weight (g), corrected for baby's sex, parity, gestational age, maternal weight and smoking habit in the multiple regression model. In comparison to the A allele, paternal G allele inheritance is associated with significantly reduced birthweight ($p = 0.01$, 95% CI −232 to −32) and head circumference ($p = 0.01$, 95% CI −0.85 to −0.11) but not with placental weight ($p = 0.98$, 95% CI −35 to 35). Paternal A/G, paternal transmission of A/G SNP rs941576; A, paternal transmission of the A allele; G, paternal transmission of the G allele; BW, birthweight; HC, head circumference; PW, placental weight. (Online version in colour.)

The rs941576 (G/A) SNP of the *DLK1-MEG3* gene region on human chromosome 14 has previously been identified as a type 1 diabetes (T1D) susceptibility locus [38]. A reduced paternal, but not maternal, transmission of the protective G allele was observed in the T1D-affected individuals, showing a clear parent-of-origin effect. It was suggested that the rs941576 variant may affect nearby paternally expressed genes, including *DLK1*. Notably, higher birthweight has been linked to increased T1D risk [66–68]. This prompted us to test whether paternal transmission of the protective G allele is associated with (i) lower *DLK1* expression and/or with (ii) reduced birthweight using the DNA samples from the Moore cohort. Because this is located within intron 6 of *MEG3* and 105 kb downstream of *DLK1*, its potential influence on *MEG3* expression was also tested.

In this study, 295 trio DNA samples from the Moore cohort were used for genotyping the rs941576 SNP. The resulting frequencies of the three genotypes were GG: 24%, AG: 45% and AA: 31%. 112 and 141 babies inherited paternal G and A bases, respectively, and 119 and 132 babies inherited maternal G and A, respectively. Using multiple linear regression analysis, we found that paternal or maternal transmission of the G allele is not correlated with *DLK1* expression ($p = 0.47$ and $p = 0.63$, respectively) or with *MEG3* expression ($p = 0.7$ and $p = 0.085$, respectively).

Next, the association between the inheritance of a paternal G allele with fetal growth was investigated, using a multiple linear regression model, adjusted for sex of the baby, parity, gestational age and maternal weight and smoking habit. Paternal transmission of the G allele was significantly associated with an average decrease of birthweight by 132 g ($p = 0.01$, 95% CI− 232 to −32; figure 3*a*), and a 0.5 cm reduction in head circumference of the baby ($p = 0.01$, 95% CI −0.85 to −0.11; figure 3*b*), but not with placental weight (−0.45 g; $p = 0.98$, 95% CI −35 to 35; figure 3*c*). Importantly, the scale of birthweight reduction (−132 g) associated with paternal G transmission is similar to that of the maternal smoking (−152 g). Maternal inheritance of the G or A allele was not associated with birthweight ($p = 0.8$), head circumference ($p = 0.62$) or placental weight ($p = 0.86$), consistent with the observed paternal effect of the protective G allele in T1D susceptibility.

(g) Other imprinted genes studied

No evidence of correlation between *H19*, *MEG3*, *PEG10*, *PEG3*, *SLC22A18*, *CDKN1C*, *PLAGL1*_imp (imprinted transcript), *PLAGL1*_all (all transcripts) or *IGFBP3* expression, in term placenta, with fetal growth was observed (summarized in table 3). In addition, we were unable to corroborate a previously reported association between *SLC22A18* expression and head circumference [31]. We did not observe any LOI in our samples, except for *PEG3*, where 2/16 (12%) samples showed biallelic expression in term placenta. Table 4 details the polymorphic variants used for each gene and imprinting analysis results.

It was not possible to test all the candidate genes in both term placenta and for CVS tissues, owing to the limited availability of material for the latter, whereas some candidates also showed a level of expression undetected by quantitative PCR. Therefore, the candidate genes have been prioritized

according to their functional relevance. Although it would have been interesting to test *GRB10* expression in CVS, head circumference measurements were not available for the KCL CVS cohort.

2. Discussion

Suboptimal or excessive intrauterine growth leads to perinatal morbidity and mortality, as well as an increased risk for adulthood diseases [4]. Finding genetic factors that regulate normal fetal growth will potentially provide more precise monitoring of intrauterine growth. Genomic imprinting epigenetically silences one parental allele resulting in monoallelic expression. It is now accepted that paternally expressed genes tend to encourage fetal growth, whereas maternally expressed genes restrict this. In this paper, the role of imprinted genes on fetal growth was explored by summarizing and connecting our previous and current findings. Although DNA methylation plays a key regulatory role in imprinted gene expression, their methylation statuses were not assessed in our samples as CVS is a limited resource. RNA expression variation is downstream of DNA methylation or other possible DNA regulatory factors, and therefore potentially more functionally relevant. An additional DNA methylation status assessment would be an interesting aspect for the future study.

(a) The early and late effectors of fetal growth

Combining past and present studies, we have investigated the correlation between fetal growth measurements and expression levels of 13 imprinted and three non-imprinted genes highly expressed in CVS tissues and term placenta (tables 1 and 2). Our candidate gene approach has identified some early and late effectors of fetal growth. We have shown that the CVS expression of *IGF2* and *IGF2R* is positively correlated to birthweight, whereas this correlation disappears in term placenta (table 3). Conversely, *PHLDA2* expression in CVS is not correlated to birthweight, whereas *PHLDA2* expression at term is strongly negatively correlated to birthweight. Although *GRB10* expression in CVS was not tested, its expression in term placenta showed a strong negative association with head circumference. These observations suggest that *IGF2* and *IGF2R* can act to set the growth potential of the baby early in the pregnancy, and two maternally expressed growth suppressing genes, *PHLDA2* and *GRB10*, act to fine tune growth in late pregnancy, potentially to avoid the risk of giving birth to a macrosomic baby. Importantly, mouse studies indicate that both *Phlda2* and *Grb10* control placental growth by mechanisms independent of *Igf2* [10,32], implying the evolution of separate pathways to control overall fetal size, possibly reflected by the difference in timing of their functional action.

The first half of placental development is characterized by a series of important trophoblast proliferation and differentiation processes, forming mature villous and extravillous structures. The second half of gestation results in an extensive vascularization and placental mass expansion [69]. Early gestational insults such as maternal diabetes have been associated with long-term effects on the fetus, owing to their influence on the initial structural formation of the placenta. It is possible that *IGF2* and *IGF2R* are key regulators of early formation of the placenta, which then sets the growth capacity of the fetus and placenta for the rest of gestation. Interestingly, overexpression of mouse *Phlda2* results in placental size reduction, with decreased glycogen storage and failed mobilization, accompanied by progressive fetal weight loss in late gestation [61]. It has been suggested that halving *Phlda2* expression by silencing the paternal allele later in gestation may promote energy provision for the fetus at this time, by increasing the glycogen stores that will be used in late gestation when there is a particularly high nutrient demand from the fetus [61].

(b) Environment and other physiological effectors on gene expression

Although placenta is fetal in origin, it is under the influence of both maternal and fetal circulation. The placental villi consist of syncytiotrophoblasts facing the maternal blood, with cytotrophoblasts in the middle and endothelial cells facing the fetal circulation [69]. Therefore, the mRNA measured in the placenta could be a result of response to the hormones and growth factors present in both maternal and fetal circulation. In this study, potential influences of environmental variations (maternal weight/BMI and maternal smoking) and physiological variation (baby's gender, gestational age and parity) on gene expression were tested.

We did not observe a correlation between maternal smoking and gene expression levels with all genes tested in both CVS and term placenta (electronic supplementary material, table S2). This result contradicts the previous report where the upregulation of placental *PHLDA2* in smokers ($n = 12$) compared with non-smokers ($n = 64$) was observed in a microarray experiment [63]. This could be due to different sensitivities between the two techniques. However, our cohorts contained more smokers ($n = 27$, Moore cohort and $n = 33$, CVS cohort; electronic supplementary material, table S1), which allows for more accurate measure of expression. *IGF2R* expression in CVS showed a positive association with maternal BMI (electronic supplementary material, figure S1*b* and table S2). This is interesting, because *IGF2R* has been found in the syncytiotrophoblast, which is in direct contact with the maternal blood circulation, and therefore possibly regulating the effect of fetal *IGF2* levels on the mother [70].

Notably, we have found a sex-biased expression of *H19* in CVS tissues, where it is expressed more highly in males (electronic supplementary material, figure S1*d*). *H19* has previously been reported to show female-biased expression in mouse eyes [71]. Therefore, *H19* expression could be dually regulated according to the sexes of the parent (imprinting) and also the baby (sexual dimorphism), in a tissue- and time-specific manner. Moreover, downregulation of *PLAGL1* in FGR placenta of females, but not males, has been reported [72]. This was not evident in our normal term placenta samples, implying FGR-specific effects. Insight into the effect of sexual dimorphism is important for understanding both normal molecular mechanisms and sex-biased disease conditions.

(c) *DLK1*, type 1 diabetes and parent-of-origin effect on fetal growth

Type 1 diabetes (T1D) is caused by autoimmune destruction of pancreatic beta cells, resulting in insulin deficiency, although its aetiology is not fully understood [68]. The *DLK1-MEG3* imprinting locus has recently been identified as a T1D susceptibility region, marked by the rs941576 SNP in which paternal inheritance of a G allele was associated with reduced risk [38].

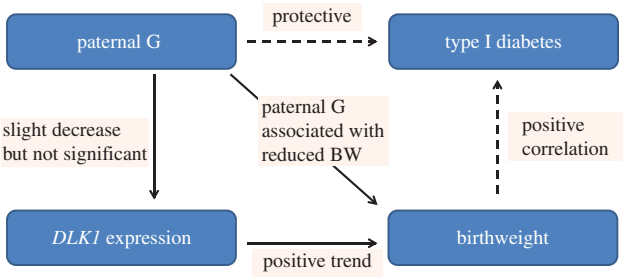


Figure 4. Current hypothesis on the association between paternal G SNP rs941576 and fetal growth. Solid lines indicate results from this study and dotted lines indicate published data [41–43]. Paternal inheritance of the G allele is associated with an average reduction in birthweight by 132 g ($p = 0.01$). The paternal G allele is also correlated with reduction in *DLK1* expression although not significantly ($p = 0.47$). There was a trend of positive association between *DLK1* expression in term placenta and birthweight ($p = 0.07$). Our hypothesis suggests that the paternal G allele reduces *DLK1* expression which causes reduction in birthweight and risk of type 1 diabetes. (Online version in colour.)

DLK1 is highly expressed in pancreatic islet cells and is involved in differentiation of pancreatic beta cells, suggesting its strong functional candidacy [73].

In this study, we found that paternal transmission of the protective G allele results in a significant decrease of birthweight, by 132 g (figure 3*a*), and head circumference, by 0.5 cm (figure 3*b*). Of note, higher birthweight has been linked to increased T1D risk [66–68]. Therefore, paternal inheritance of the G allele may give protective effect from T1D via its association with reduced birthweight. This could also be associated with a decrease in *DLK1* expression although this association did not reach statistical significance. Importantly, the magnitude of birthweight reduction (−132 g) and head circumference (−0.5 cm) related to the paternal G allele inheritance was similar to that observed for the increase in birthweight (+155 g) and in head circumference (+0.23 cm) caused by inheriting a *PHLDA2* promoter RS1 allele from a RS1 homozygous mother [36]. Our current working hypothesis regarding the relationship between the role of *DLK1* in fetal growth and T1D is described in figure 4 [37,64,65].

3. Conclusion

We have identified that expression of *IGF2* and *IGF2R* in early placenta (CVS) are positively correlated to CRL and birthweight, but not in term placenta when the oppositely maternally expressed genes *PHLDA2* and *GRB10* act to negatively regulate growth. We have also identified that the paternal transmission of the T1D protective G allele of rs941576 SNP results in a significant reduction in birthweight ($p = 0.01$, 95% CI −232 to −32), emphasizing the importance of accounting for parent-of-origin effects when analysing genomic data. Characterization of genes important in intrauterine growth will allow a more accurate surveillance of fetal growth and help identify targets for clinical intervention in suboptimal pregnancies. During pregnancy, a combination of different levels of imprinted genes or genetic predispositions will affect the baby's birthweight. An additional environmental layer is added by maternal smoking. Further investigation of all these candidates is warranted in larger cohorts to identify further genetic variants that exhibit parent-of-origin associated growth regulation and to find gene expression variations. Together with previously known genetic variants associated with fetal growth (reviewed in [26]), and expression studies, these may be used as an effective, combined diagnostic tool to identify and predict growth-restricted and macrosomic babies, which would provide huge benefits for the short- and long-term health of both mother and baby.

4. Materials

(a) King's College London chorionic villus sample cohort

CVS was carried out between 11 and 13 weeks of gestation in 355 singleton pregnancies that were followed by normal live birth at term. Participants were undergoing CVS for prenatal diagnosis for chromosomal abnormality at King's College Hospital London. The samples used in this study were obtained from excess CVS tissues from fully ethically consented women, and the research was approved by the King's College Hospital Ethics Committee. The medical records of this cohort are summarized in the electronic supplementary material, table S1 [17].

(b) Moore cohort

The Moore cohort consists of 302 consented white European trios recruited at Queen Charlotte's and Chelsea Hospital between 2003 and 2004 [33]. The placental samples were collected from ultrasound dated, live birth singleton pregnancies. Each placental sample was dissected into four pieces near the umbilical cord insertion point, washed in phosphate-buffered saline, snap-frozen in liquid nitrogen and stored at −80°C. Parental blood samples (10 ml) were collected in EDTA tubes. The medical records and characteristics of the Moore cohort are summarized in the electronic supplementary, table S1.

5. Methods

(a) DNA and RNA extraction

Total RNA from term placental tissue was extracted using Trizol reagent (Life Technologies), and treated with TURBO DNase (Ambion) according to the manufacturers guidelines. Fetal DNA from 1 g of term placental tissue and parental DNA from 2 ml of whole blood were isolated using a standard phenol–chloroform protocol. RNA and DNA from CVS tissues were extracted by the iPrep PureLink total RNA and TrizolPlus RNA kit, including the DNase treatment and iPrep Charge-Switch gDNA tissue kit using the iPrep purification instrument (Life Technologies) following the manufacturer's instructions. The quantity and purity of nucleic acid was measured by NanoDrop ND-1000 spectrophotometer (Thermo Scientific). Only RNA samples with the 260/260 ratio in the range of 2 ± 0.2 were used for further study.

(b) Reverse transcription

A first strand of complementary DNA (cDNA) was synthesized from 1 µg (term placenta) and 100 ng (CVS) of RNA with Moloney murine leukemia virus reverse transcriptase (M-MLV RT) according to the manufacturer's instructions (Promega). Duplicate sets of samples without reverse transcriptase were made as negative controls to detect any genomic contamination in RNA samples. The conversion of RNA to cDNA was confirmed by polymerase chain reaction (PCR) with Taq DNA polymerase

(Bioline) beta-actin (*ACTB*) primers (electronic supplementary material table S5).

(c) Quantitative polymerase chain reaction

qPCR was performed using the *Power* SYBRGreen PCR master mix (Life Technologies). Each sample was tested in triplicate, and each plate contained a no-template-control and a cDNA pool as a reference sample to control for interplate variations. The reaction plate was placed on the StepOne plus real-time PCR systems, analysed in the comparative C_t mode. Ribosomal protein L19 (*L19*) housekeeping gene was used as an endogenous control throughout the experiments. Thermal cycle conditions consist of initial incubation at 50°C for 2 min for one cycle, polymerase activation at 95°C for 10 min for one cycle and 40 cycles of denaturation at 95°C for 15 s, and annealing and extension at 60°C for 1 min. The efficiency of the primers was determined by running a standard curve and calculated by $10^{(-1/\text{slope})} - 1$. The qPCR primer sequences are provided in the electronic supplementary material, table S4. The resulting data were analysed with the StepOne v. 2.1 software to obtain relative quantification (RQ) values, using the formula $RQ = 2^{-\Delta\Delta Ct}$.

(d) Imprinting analysis

Monoallelic expression of genes was investigated by sequencing gene-specific amplicons from cDNA samples that corresponded to genomic DNA heterozygous for selected SNPs. Parental DNA was available for term placental samples, and was used for sequencing to check the parental origin of the expressed allele. SNPs with relatively high average heterozygosity were chosen for each gene within the exon covering all isoforms. PCR primer sequences are summarized in the electronic supplementary material, table S5, and the list of selected SNPs is found in table 4. Sequencing was carried out using the BigDye terminator v. 1.1 cycle sequencing kit (Life Technologies), and the read-out was analysed with Sequencher v. 4.8 (Gene Codes Corporation).

(e) Statistical analysis

All statistical analyses were performed using the R software (R Foundation for Statistical Computing). The relative expression of the candidate genes in term placenta was correlated to the baby's birth weight, placental weight and head circumference using a multiple linear regression model adjusted for baby's sex, gestational age, parity, maternal weight/BMI and smoking habits. These variables used in the model have previously been established as confounding factors in our previous studies in the same cohort [33,36]. A logarithmic scale was used for the expression values when appropriate, and BIC test was performed to check the fit of the models. A significance threshold of 5% was used in the analysis.

Ethics statement. The study was approved by the Hammersmith and Queen Charlotte's and Chelsea Hospitals' Trust Research Ethics Committee (registration no. 2001/6029).

Acknowledgements. We thank Drs Helen Stevens and Chris Wallace for genotyping data and performing initial statistical analysis on the Moore cohort DNA. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting the families. G.E.M. co-wrote the manuscript, contributed the concept and design of the experiments, and provided intellectual contributions. M.I. co-wrote the manuscript, performed RT-qPCR and sequencing experiments, collated the experimental data and performed statistical analysis. C.D., L.A., L.J.L, J.L.S., Y.C., R.B., M.B., I.I., O.O., R.A. and D.M. prepared RNA samples, performed RT-qPCR and sequencing experiments and analysed the data. A.J.D. prepared RNA samples and performed RT-qPCR experiments and analysed the data. S.A. and A.C.T. prepared RNA and DNA samples, performed RT-qPCR and sequencing experiments and analysed the data. S.A.A. prepared RNA and DNA samples and provided intellectual contributions. J.M.F. prepared RNA and DNA samples and designed RT-PCR primers. L.R. provided obstetric advice and discussion. A.S. helped collation of the CVS data. K.H.N. provided the CVS samples. J.C.W. performed statistical analysis. P.S. provided intellectual contributions and revised the manuscript.

Funding statement. G.E.M. Fetal growth and development research team is funded by the MRC, Wellbeing of Women, March of Dimes, SPARKS, Wellcome Trust and the Great Ormond Street Hospital Children's Charity (GOSHCC). M.I. was funded for her PhD by the Child Health Research Appeal Trust (the Institute of Child Health and the Great Ormond Street Hospital for Children), Overseas Research Studentship, the Medical Research Council (MRC). C.D. is funded by Save the Baby Unit, P.S. is supported by GOSHCC.

Competing interests. We have no competing interests.

References

1. Basso O, Wilcox AJ, Weinberg CR. 2006 Birth weight and mortality: causality or confounding? *Am. J. Epidemiol.* **164**, 303–311. (doi:10.1093/aje/kwj237)
2. Brodsky D, Christou H. 2004 Current concepts in intrauterine growth restriction. *J. Intensive Care Med.* **19**, 307–319. (doi:10.1177/0885066604269663)
3. Shankar M, Navti O, Amu O, Konje JC. 2002 Assessment of stillbirth risk and associated risk factors in a tertiary hospital. *J. Obstet. Gynaecol.* **22**, 34–38. (doi:10.1080/01443610120101682)
4. Simmons RA. 2009 Developmental origins of adult disease. *Pediat. Clin. N. Am.* **56**, 449–466. (doi:10.1016/j.pcl.2009.03.004)
5. Abu-Amero S, Monk D, Apostolidou S, Stanier P, Moore G. 2006 Imprinted genes and their role in human fetal growth. *Cytogenet. Genome Res.* **113**, 262–270. (doi:10.1159/000090841)
6. Barton SC, Surani MA, Norris ML. 1984 Role of paternal and maternal genomes in mouse development. *Nature* **311**, 374–376. (doi:10.1038/311374a0)
7. McGrath J, Solter D. 1984 Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**, 179–183. (doi:10.1016/0092-8674(84)90313-1)
8. Surani MA, Barton SC, Norris ML. 1984 Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548–550. (doi:10.1038/308548a0)
9. Reik W, Dean W, Walter J. 2001 Epigenetic reprogramming in mammalian development. *Science* **293**, 1089–1093. (doi:10.1126/science.1063443)
10. Charalambous M, Smith FM, Bennett WR, Crew TE, Mackenzie F, Ward A. 2003 Disruption of the imprinted Grb10 gene leads to disproportionate overgrowth by an Igf2-independent mechanism. *Proc. Natl Acad. Sci. USA* **100**, 8292–8297. (doi:10.1073/pnas.1532175100)
11. Lau MM, Stewart CE, Liu Z, Bhatt H, Rotwein P, Stewart CL. 1994 Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes Dev.* **8**, 2953–2963. (doi:10.1101/gad.8.24.2953)
12. Lefebvre L, Viville S, Barton SC, Ishino F, Keverne EB, Surani MA. 1998 Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest. *Nat. Genet.* **20**, 163–169. (doi:10.1038/2464)
13. Leighton PA, Ingram RS, Eggenschwiler J, Efstratiadis A, Tilghman SM. 1995 Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature* **375**, 34–39. (doi:10.1038/375034a0)
14. Li L, Keverne EB, Aparicio SA, Ishino F, Barton SC, Surani MA. 1999 Regulation of maternal behavior and offspring growth by paternally expressed Peg3. *Science* **284**, 330–333. (doi:10.1126/science.284.5412.330)

15. Varrault A *et al*. 2006 Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. *Dev. Cell.* **11**, 711–722. (doi:10.1016/j.devcel.2006.09.003)
16. Mackay DJ, Temple IK. 2010 Transient neonatal diabetes mellitus type 1. *Am. J. Med. Genet. C Semin. Med. Genet.* **154C**, 335–342. (doi:10.1002/ajmg.c.30272)
17. Demetriou C *et al*. 2014 Paternally expressed, imprinted insulin-like growth factor-2 in chorionic villi correlates significantly with birth weight. *PLoS ONE* **9**, e85454. (doi:10.1371/journal.pone.0085454)
18. Abu-Amero S, Monk D, Frost J, Preece M, Stanier P, Moore GE. 2008 The genetic aetiology of Silver–Russell syndrome. *J. Med. Genet.* **45**, 193–199. (doi:10.1136/jmg.2007.053017)
19. Ono R *et al*. 2006 Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* **38**, 101–106. (doi:10.1038/ng1699)
20. Lim AL *et al*. 2012 Epigenetic state and expression of imprinted genes in umbilical cord correlates with growth parameters in human pregnancy. *J. Med. Genet.* **49**, 689–697. (doi:10.1136/jmedgenet-2012-100858)
21. Diplas AI, Lambertini L, Lee MJ, Sperling R, Lee YL, Wetmur J, Chen J. 2009 Differential expression of imprinted genes in normal and IUGR human placentas. *Epigenetics* **4**, 235–240. (doi:10.4161/epi.9019)
22. Hannula K, Lipsanen-Nyman M, Kontiokari T, Kere J. 2001 A narrow segment of maternal uniparental disomy of chromosome 7q31-qter in Silver–Russell syndrome delimits a candidate gene region. *Am. J. Hum. Genet.* **68**, 247–253. (doi:10.1086/316937)
23. Esquiliano DR, Guo W, Liang L, Dikkes P, Lopez MF. 2009 Placental glycogen stores are increased in mice with H19 null mutations but not in those with insulin or IGF type 1 receptor mutations. *Placenta* **30**, 693–699. (doi:10.1016/j.placenta.2009.05.004)
24. Gicquel C *et al*. 2005 Epimutation of the telomeric imprinting center region on chromosome 11p15 in Silver–Russell syndrome. *Nat. Genet.* **37**, 1003–1007. (doi:10.1038/ng1629)
25. Begemann M, Spengler S, Gogiel M, Grasshoff U, Bonin M, Betz RC, Dufke A, Spier I, Eggermann T. 2012 Clinical significance of copy number variations in the 11p15.5 imprinting control regions: new cases and review of the literature. *J. Med. Genet.* **49**, 547–553. (doi:10.1136/jmedgenet-2012-100967)
26. Ishida M, Moore GE. 2013 The role of imprinted genes in humans. *Mol. Aspects Med.* **34**, 826–840. (doi:10.1016/j.mam.2012.06.009)
27. Tunster SJ, Van de Pette M, John RM. 2011 Fetal overgrowth in the Cdkn1c mouse model of Beckwith–Wiedemann syndrome. *Dis. Models Mech.* **4**, 814–821. (doi:10.1242/dmm.007328)
28. Arboleda VA *et al*. 2012 Mutations in the PCNA-binding domain of CDKN1C cause IMAGe syndrome. *Nat. Genet.* **44**, 788–792. (doi:10.1038/ng.2275)
29. Choufani S, Shuman C, Weksberg R. 2010 Beckwith–Wiedemann syndrome. *Am. J. Med. Genet. C Semin. Med. Genet.* **154C**, 343–354. (doi:10.1002/ajmg.c.30267)
30. Brioude F *et al*. 2013 CDKN1C mutation affecting the PCNA-binding domain as a cause of familial Russell Silver syndrome. *J. Med. Genet.* **50**, 823–830. (doi:10.1136/jmedgenet-2013-101691)
31. Lambertini L, Marsit CJ, Sharma P, Maccani M, Ma Y, Hu J, Chen J. 2012 Imprinted gene expression in fetal growth and development. *Placenta* **33**, 480–486. (doi:10.1016/j.placenta.2012.03.001)
32. Frank D *et al*. 2002 Placental overgrowth in mice lacking the imprinted gene Ipl. *Proc. Natl Acad. Sci. USA* **99**, 7490–7495. (doi:10.1073/pnas.122039999)
33. Apostolidou S *et al*. 2007 Elevated placental expression of the imprinted PHLDA2 gene is associated with low birth weight. *J. Mol. Med. (Berl.)* **85**, 379–387. (doi:10.1007/s00109-006-0131-8)
34. Kumar N, Leverence J, Bick D, Sampath V. 2012 Ontogeny of growth-regulating genes in the placenta. *Placenta* **33**, 94–99. (doi:10.1016/j.placenta.2011.11.018)
35. McMinn J, Wei M, Schupf N, Cusmai J, Johnson EB, Smith AC, Weksberg R, Thaker HM, Tycko B. 2006 Unbalanced placental expression of imprinted genes in human intrauterine growth restriction. *Placenta* **27**, 540–549. (doi:10.1016/j.placenta.2005.07.004)
36. Ishida M *et al*. 2012 Maternal inheritance of a promoter variant in the imprinted PHLDA2 gene significantly increases birth weight. *Am. J. Hum. Genet.* **90**, 715–719. (doi:10.1016/j.ajhg.2012.02.021)
37. Moon YS, Smas CM, Lee K, Villena JA, Kim KH, Yun EJ, Sul HS. 2002 Mice lacking paternally expressed Pref-1/Dlk1 display growth retardation and accelerated adiposity. *Mol. Cell. Biol.* **22**, 5585–5592. (doi:10.1128/MCB.22.15.5585-5592.2002)
38. Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, Clayton DG. 2010 The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nat. Genet.* **42**, 68–71. (doi:10.1038/ng.493)
39. Takahashi N, Okamoto A, Kobayashi R, Shirai M, Obata Y, Ogawa H, Sotomaru Y, Kono T. 2009 Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. *Hum. Mol. Genet.* **18**, 1879–1888. (doi:10.1093/hmg/ddp108)
40. Feng W, Marquez RT, Lu Z, Liu J, Lu KH, Issa JP, Fishman DM, Yu Y, Bast RC. 2008 Imprinted tumor suppressor genes ARHI and PEG3 are the most frequently down-regulated in human ovarian cancers by loss of heterozygosity and promoter methylation. *Cancer* **112**, 1489–1502. (doi:10.1002/cncr.23323)
41. Moore T, Haig D. 1991 Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet.* **7**, 45–49. (doi:10.1016/0168-9525(91)90230-N)
42. Court F *et al*. 2014 Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome Res.* **24**, 554–569. (doi:10.1101/gr.164913.113)
43. Lofqvist C *et al*. 2007 IGFBP3 suppresses retinopathy through suppression of oxygen-induced vessel loss and promotion of vascular regrowth. *Proc. Natl Acad. Sci. USA* **104**, 10 589–10 594. (doi:10.1073/pnas.0702031104)
44. Deal C *et al*. 2001 Novel promoter polymorphism in insulin-like growth factor-binding protein-3: correlation with serum levels and interaction with known regulators. *J. Clin. Endocrinol. Metab.* **86**, 1274–1280.
45. Baker J, Liu JP, Robertson EJ, Efstratiadis A. 1993 Role of insulin-like growth factors in embryonic and postnatal growth. *Cell* **75**, 73–82. (doi:10.1016/0092-8674(93)90680-O)
46. Netchine I, Azzi S, Le Bouc Y, Savage MO. 2011 IGF1 molecular anomalies demonstrate its critical role in fetal, postnatal growth and brain development. *Best Pract. Res. Clin. Endocrinol. Metab.* **25**, 181–190. (doi:10.1016/j.beem.2010.08.005)
47. Gannagé-Yared M-H, Klammt J, Chouery E, Corbani S, Mégarbané H, Ghoch JA, Choucair N, Pfäffle R, Mégarbané A. 2013 Homozygous mutation of the IGF1 receptor gene in a patient with severe pre- and postnatal growth failure and congenital malformations. *Eur. J. Endocrinol.* **168**, K1–K7. (doi:10.1530/EJE-12-0701)
48. Hiden U, Glitzner E, Hartmann M, Desoye G. 2009 Insulin and the IGF system in the human placenta of normal and diabetic pregnancies. *J. Anat.* **215**, 60–68. (doi:10.1111/j.1469-7580.2008.01035.x)
49. Frasca F *et al*. 1999 Insulin receptor isoform A, a newly recognized, high-affinity insulin-like growth factor II receptor in fetal and cancer cells. *Mol. Cell. Biol.* **19**, 3278–3288.
50. Monk D, Arnaud P, Apostolidou S, Hills FA, Kelsey G, Stanier P, Feil R, Moore GE. 2006 Limited evolutionary conservation of imprinting in the human placenta. *Proc. Natl Acad. Sci. USA* **103**, 6623–6628. (doi:10.1073/pnas.0511031103)
51. Willison K. 1991 Opposite imprinting of the mouse Igf2 and Igf2r genes. *Trends Genet.* **7**, 107–109. (doi:10.1016/0168-9525(91)90441-R)
52. Vu TH, Li T, Nguyen D, Nguyen BT, Yao XM, Hu JF, Hoffman AR. 2000 Symmetric and asymmetric DNA methylation in the human IGF2-H19 imprinted region. *Genomics* **64**, 132–143. (doi:10.1006/geno.1999.6094)
53. Bukowski R *et al*. 2007 Human sexual size dimorphism in early pregnancy. *Am. J. Epidemiol.* **165**, 1216–1218. (doi:10.1093/aje/kwm024)
54. Holt LJ, Siddle K. 2005 Grb10 and Grb14: enigmatic regulators of insulin action–and more? *Biochem. J.* **388**, 393–406. (doi:10.1042/BJ20050216)
55. Blagitko N, Mergenthaler S, Schulz U, Wollmann HA, Craigen W, Eggermann T, Ropers H-H, Kalscheuer VM. 2000 Human GRB10 is imprinted and expressed from the paternal and maternal

allele in a highly tissue- and isoform-specific fashion. *Hum. Mol. Genet.* **9**, 1587–1595. (doi:10.1093/hmg/9.11.1587)

56. Monk D, Arnaud P, Frost J, Hills FA, Stanier P, Feil R, Moore GE. 2009 Reciprocal imprinting of human GRB10 in placental trophoblast and brain: evolutionary conservation of reversed allelic expression. *Hum. Mol. Genet.* **18**, 3066–3074. (doi:10.1093/hmg/ddp248)
57. Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C. 2010 High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* **329**, 643–648. (doi:10.1126/science.1190830)
58. Garfield AS *et al*. 2011 Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature* **469**, 534–538. (doi:10.1038/nature09651)
59. Saxena A, Morozov P, Frank D, Musalo R, Lemmon MA, Skolnik EY, Tycko B. 2002 Phosphoinositide binding by the pleckstrin homology domains of Ipl and Tih1. *J. Biol. Chem.* **277**, 49935–49944. (doi:10.1074/jbc.M206497200)
60. Salas M, John R, Saxena A, Barton S, Frank D, Fitzpatrick G, Higgins MJ, Tycko B. 2004 Placental growth retardation due to loss of imprinting of Phlda2. *Mech. Dev.* **121**, 1199–1210. (doi:10.1016/j.mod.2004.05.017)
61. Tunster SJ, Tycko B, John RM. 2010 The imprinted Phlda2 gene regulates extraembryonic energy stores. *Mol. Cell. Biol.* **30**, 295–306. (doi:10.1128/MCB.00662-09)
62. Doria S, Sousa M, Fernandes S, Ramalho C, Brandao O, Matias A, Barros A, Carvalho F. 2010 Gene expression pattern of IGF2, PHLDA2, PEG10 and CDKN1C imprinted genes in spontaneous miscarriages or fetal deaths. *Epigenetics* **5**, 444–450. (doi:10.4161/epi.5.5.12118)
63. Bruchova H, Vasikova A, Merkerova M, Milcova A, Topinka J, Balascak I, Pastorkova A, Sram RJ, Brdicka R. 2010 Effect of maternal tobacco smoke exposure on the placental transcriptome. *Placenta* **31**, 186–191. (doi:10.1016/j.placenta.2009.12.016)
64. Laborda J, Sausville EA, Hoffman T, Notario V. 1993 Dlk, a putative mammalian homeotic gene differentially expressed in small cell lung carcinoma and neuroendocrine tumor cell line. *J. Biol. Chem.* **268**, 3817–3820.
65. Sul HS. 2009 Minireview: Pref-1: role in adipogenesis and mesenchymal cell fate. *Mol. Endocrinol.* **23**, 1717–1725. (doi:10.1210/me.2009-0160)
66. Cardwell CR *et al*. 2010 Birthweight and the risk of childhood-onset type 1 diabetes: a meta-analysis of observational studies using individual patient data. *Diabetologia* **53**, 641–651. (doi:10.1007/s00125-009-1648-5)
67. Harder T, Roepke K, Diller N, Stechling Y, Dudenhausen JW, Plagemann A. 2009 Birth weight, early weight gain, and subsequent risk of type 1 diabetes: systematic review and meta-analysis. *Am. J. Epidemiol.* **169**, 1428–1436. (doi:10.1093/aje/kwp065)
68. Stene LC, Magnus P, Lie RT, Sovik O, Joner G. 2001 Birth weight and childhood onset type 1 diabetes: population based cohort study. *BMJ* **322**, 889–892. (doi:10.1136/bmj.322.7291.889)
69. Desoye G, Hauguel-de Mouzon S. 2007 The human placenta in gestational diabetes mellitus. The insulin and cytokine network. *Diabetes Care.* **30**(Suppl. 2), S120–S126. (doi:10.2337/dc07-s203)
70. Harris LK, Crocker IP, Baker PN, Aplin JD, Westwood M. 2011 IGF2 actions on trophoblast in human placenta are regulated by the insulin-like growth factor 2 receptor, which can function as both a signaling and clearance receptor. *Biol. Reprod.* **84**, 440–446. (doi:10.1095/biolreprod.110.088195)
71. Reinius B, Kanduri C. 2013 Elevated expression of H19 and Igf2 in the female mouse eye. *PLoS ONE* **8**, e56611. (doi:10.1371/journal.pone.0056611)
72. Iglesias-Platas I, Martin-Trujillo A, Petazzi P, Guillaumet-Adkins A, Esteller M, Monk D. 2014 Altered expression of the imprinted transcription factor *PLAGL1* deregulates a network of genes in the human IUGR placenta. *Hum. Mol. Genet.* **23**, 6275–6285. (doi:10.1093/hmg/ddu347)
73. Laborda J. 2000 The role of the epidermal growth factor-like protein dlk in cell differentiation. *Histol. Histopathol.* **15**, 119–129.

General Practice

Abu-Amero et al., JGPR 2014, 2:3
http://dx.doi.org/10.4172/2329-9126.1000158

Review Article Open Access

The Baby Bio Bank-A Legacy for Researchers Worldwide into Common Complications of Pregnancy

Sayeda Abu-Amero[1*], Anna Thomas[1], Shawnelle White[2], Katherine Rogers[2], Ana Maria Perez Miranda[3], Nita Solanky[4], Lydia Leon[1], Charalambos Demetriou[1], Xiayi Ke[1], Sam Stanier[2], Ben Stanier[2], Mr Harry Costello[2] Miss Samrawit Tzehaie[2], Ms Lara Al-Olabi[2], Catherine Williamson[3], Mark Johnson[4], Lesley Regan[2*], and Gudrun E Moore[1*]

[1]co-Directors of the BBB aInstitute of Child Health, 30 Guilford Street, London, United Kingdom

[2]St Mary's Hospital, Praed Street, London, United Kingdom

[3]Institute of Reproductive and Developmental Biology, Imperial College London, Du Cane Road, London, United Kingdom

[4]Chelsea and Westminster Hospital, 369, Fulham Road, London, United Kingdom

[*]Corresponding author: Sayeda Abu-Amero, Institute of Child Health, 30 Guilford Street, London, United Kingdom, Tel : +44 207 905 2868, Email : s.abu-amero@ucl.ac.uk

Received date: Apr 04, 2014; **Accepted date:** Apr 14, 2014; **Published date:** Apr 20, 2014

Abstract

Two London Universities, University College London and Imperial College London, have established a biobank as a resource for investigating the four main complications of pregnancy (recurrent miscarriage, preterm birth, fetal growth restriction (FGR) and pre-eclampsia), that collectively affect about 20% of families trying to conceive. Samples are being taken from the three key members of the family, mother, father and baby, allowing hereditary factors from both parents to be tracked. Additional samples are being stored to allow parallel analysis of the functional mechanisms of pregnancy including the epigenetic, anatomical, physiological and even metabolic causes for the high loss of fecundity in man. http://www.ucl.ac.uk/babybiobank

Introduction

Collections of human biological samples from large cohorts of families with carefully collected phenotypic and clinical data allows detailed study of the mechanisms behind many diseases. For many years researchers have used animal models to understand the effects of genetic mutations on phenotype, but have often acknowledged that to translate knowledge fully to human disorders they must have access to corresponding matched human samples. There are some ethical barriers to overcome with humans but access to tissue samples post-surgery, a small extra blood sample during routine procedures and buccal/saliva swabs have provided invaluable material for human research. The Baby Bio Bank is focussing on the four main complications of pregnancy, recurrent miscarriage, preterm birth, FGR and pre-eclampsia.

The genetic and environmental mechanisms of growth and development in utero are complex as both genes and the environment have interacting effects. There have been many studies showing gene associations with complications of pregnancy, but due to small numbers of samples it is often difficult to draw definitive conclusions. The Baby Bio Bank (BBB) will provide the necessary number of samples for scientific research into complications of pregnancy in order to identify the genes and biological mechanisms involved. By creating a bioresource for national and international research, the BBB aims to provide samples of high quality so that data can be universally compared. It will also foster collaborative research to ensure the best use is made of the samples. By providing these samples as a biobank, the burden in terms of ethics applications, staff costs and recruitment for individual projects can be minimised.

Biobanking

Biobanking is defined as the collection and storage of samples and can vary from the collection of a single sample to millions, for individual use to global use and from simple generic to highly specialized collections. In most cases biobanks are established for the purpose of health research and translational medicine. This is the only sensible way forward for the study of complex, multifactorial diseases in humans where hundreds of samples are needed to elucidate function.

The UK hosts several large biobanks. The UK Biobank (http://www.ukbiobank.ac.uk) is perhaps best known, and its purpose is for specific research into diseases occurring in mid- to late life such as cancer, heart disease, diabetes, dementia, depression, osteoporosis, arthritis, Parkinson's disease and lung and kidney disorders. It has now recruited 500,000 individuals since its beginning in 2006, with 22 assessment centres around the country The UK Biobank collects and stores blood, urine and saliva with an extensive set of medical data on blood pressure, lung function and grip strength, height, weight and body mass, arterial stiffness, vision, hearing, family history of common diseases, bone density, diet and fitness.

Biobanks can also complement one another, such as the The Confederation of Cancer Biobanks (CCB-http://www.oncoreuk.org) which is a consortium of UK based organisations involved in the development, management and use of biobank resources for cancer research. The CCB also promotes the transfer of knowledge and experience amongst smaller biobanks to promote collaboration rather than competition. Other biobanks are more specialised, for example the Oxford Pregnancy Biobank, which collects maternal blood and urine, as well as clinical data such as scans, specifically for research

into micro- and nanoparticle biology in the search for new biomarkers in pre-eclampsia (http://www.obs-gyn.ox.ac.uk/research/ian-sargent).

The Baby Bio Bank

The BBB is directed by Professors Gudrun Moore (Institute of Child Health) and Lesley Regan (St Mary's Hospital, Paddington) supported by two clinical collaborators, Professors Catherine Williamson (Queen Charlotte's and Chelsea Hospital), Mark Johnson (Chelsea and Westminster Hospital), The BBB manager, Dr Sayeda Abu-Amero is responsible for all administrative aspects including HTA compliance, ethics application, annual reports, website management, biannual newsletter and management of the recruitment and sample management team. There are three recruiters (Ms Katherine Rogers, Drs Ana Maria Perez Miranda and Nita Solanky) working at the different hospital sites and they provide 24 hour cover to ensure as many samples are collected to complete trios as well as two sample management personnel (Drs Shawnelle White and Anna Thomas) responsible for sample barcoding, processing, quality control, storage as well as database entry.

The BBB has its own Research Management Board (RMB) currently chaired by Dame Joan Higgins which is responsible for ensuring all protocols are carried out according to national tissue bank ethics and the HTA and will also be responsible for determining which projects are successful at obtaining samples for future research.

Recurrent miscarriage is the loss of 3 or more consecutive pregnancies, and affects 1-2% of couples trying to have a successful pregnancy outcome [1]. While the actual number is difficult to quantify because of the nature of this complication, experts estimate there are at least 6,000 couples newly affected every year in the UK (according to the American College of Obstetrician and Gynaecologists) and miscarriages are the most common type of pregnancy loss (http://www.nhs.uk/conditions/miscarriage/Pages/Introduction.aspx).

One in every 13 babies is born before 37 weeks and is classed as a preterm birth. Prematurity is responsible for 80% of all neonatal deaths, and affects 30,000 babies in the UK annually. The rate of cerebral palsy in preterm babies is up to 30 times higher than in babies born at term. This incidence has not changed significantly in recent years, and prematurity represents a major health issue and challenge for modern obstetric care (http://www.nhs.uk/planners/pregnancycareplanner/pages/prematurelabour.aspx) [2].

Fetal Growth Restriction (FGR previously referred to as IUGR (intra-uterine growth restriction)) refers to a condition in which a fetus is unable to achieve its genetically determined potential size and affects 20,000 babies in the UK every year and this condition accounts for 50% of all stillbirths. Defining true fetal growth restriction (for example, by identifying falling intrauterine growth profiles) is difficult and defining a group of babies at or below the 10th percentile is often used as a proxy. Although 60% of these will turn out to be normal small babies, the remainder are at an increased risk of potentially preventable perinatal death. (http://www.patient.co.uk/doctor/Intrauterine-Growth-Retardation.htm) [3].

Pre-eclampsia is a multisystem disorder that typically presents with high blood pressure and proteinuria in the second half of pregnancy and is a leading cause of maternal death. Mild pre-eclampsia affects 4-6% of first time pregnancies with 1% of pregnant women experiencing severe pre-eclampsia [4]. Women who have pre-eclampsia in one pregnancy are at a higher risk of developing the condition in subsequent pregnancies. It is responsible for a considerable proportion of the 500,000 infant deaths per year worldwide and around six women in the UK die annually as a result of complications associated with pre-eclampsia (http://www.nhs.uk/Conditions/Pre-eclampsia/Pages/Introduction.aspx).

There is evidence that the four complications of pregnancy are part of a biological continuum which originates at the time of embryonic implantation with interrelated genetic and environmental factors influencing the final outcome of the pregnancy (Figure 1). Using samples collected by the BBB and advanced scientific technology capable of analysing many samples in parallel we can begin to dissect the biological, physiological, genetic and even epigenetic causes of these complications. The results should give rise to predictive biomarkers that might identify those most at risk prior to disease onset. This should offer the opportunity to design therapies to prevent disease or reduce the severity of adverse maternal and fetal outcomes associated with recurrent miscarriage, preterm birth, FGR and pre-eclampsia.

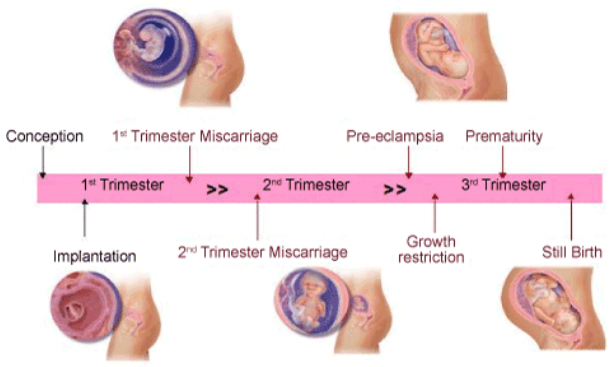


Figure 1: The four complications of pregnancy targeted by the BBB are a continuum.

The BBB collection will provide the biological materials for research into the causes of these and related pregnancy complications. The target of the BBB is to recruit 500 white European trios (mother, father and baby) from pregnancies affected by each complication plus 500 normal trios for comparative purposes giving a total of 2500 trios over the course of five years. Mothers and their partners are recruited in a variety of clinics but most often at their first booking in the antenatal clinic. Blood samples for DNA, serum and plasma, are collected from all consenting participants, plus urine from the mother. Where father is reluctant to give blood we the option of a saliva samples which provides high yields of excellent quality DNA [5]. The families are tracked throughout their pregnancy, and at the birth of the baby, we are notified and collect pieces of placenta, umbilical cord and cord blood. Where the baby is missed because delivery is at another hospital or any other reason, recruiters aim to get a buccal swab. Where no fetal sample is available-we are still able to use the maternal and/or paternal samples with available clinical information for other studies not looking at inherited genetic markers.

Ethical approval for the BBB was a straight forward process using the National Research Ethics Service (NRES) via the online Integrated Research Application System (IRAS-https://www.myresearchproject.org.uk/) and communications with Trent

Research Ethics Committee (REC) (new NRES committee name is East Midlands Derby 1 REC) have been efficient. We applied for Research Tissue Bank (RTB-https://www.myresearchproject.org.uk/help/ResearchTissueBanks.aspx) ethics which is voluntary and has the benefit that 'favourable ethical opinion applies to all research projects conducted in the UK using tissue or data supplied by the tissue bank, provided that the release of tissue or data complies with the attached conditions' (see https://www.myresearchproject.org.uk/help/Help%20Documents/PdfDocuments/researchtissuebank.pdf). This means that the BBB RMG are able to apply ethical approval to peer reviewed projects which relieves the potential research groups from the time consuming process of ethics application for their studies. Interested researchers are able to apply via the WoW BBB website (http://www.wellbeingofwomen.org.uk/research/baby-bio-bank/?menu=3c) where they submit a brief description of the proposal, type, number and amount of samples required, proof of funding and local R&D peer review. Applicants are also able to apply at the same time as applying for grants to help speed the research process. Applications are acknowledged by the BBB Manager and submitted to the BBB RMB for consideration every 4 months. Applicants are informed within 1 month of the BBB RMB meeting if their application is successful and samples dispatched within 3 months of approval. All applicants will have to submit an annual progress report and any unused samples must be returned to the BBB. The WoW BBB RMB recently met in March 2014 and approved the first three projects from within the UK-one with funding and two pending funding.

The hospitals in the BBB portfolio were selected as representing clinics with high proportions of pregnancies with complications. In addition, many research projects and clinical trials are also conducted in these (and other) hospitals for the ultimate benefit of the participant. The BBB recruiting staff put the wellbeing and satisfaction of the participant at the top of their list when considering recruitment as this is fundamental to the success of any biobank or research project involving patients [6].

BBB recruiters spending some time discussing the patient information sheet with the couple, giving the couple time to go home and discuss any issues if necessary before consenting and arranging for the bloods to be taken at the participant's convenience.

As all the samples are intended for DNA, RNA and protein isolation, they are being collected, processed and stored to the highest possible scientific standards. Quality control audits followed by downstream applications such as PCR, real-time PCR and sequencing has shown clearly that the samples are of high quality and can be used with confidence by researchers. The biological specimens have restricted value without the relevant clinical information for the pregnancy under investigation. We have ethical approval to include clinical information from the hospital databases relating to factors affecting pregnancy such as parental height, weight and relevant medical history such as diabetes, hypertension, and smoking. Importantly, we also collect fetal outcome data such as gestational age, birth weight, placental weight and mode of delivery. Consent is obtained from parents for access to their notes and information stored in the hospital database and also for any relevant queries which may be missing/unavailable to be followed up with the participants themselves. Currently, the BBB database stores detailed sample information (type of sample, amount or volumes available, date collected, time to process sample) in APOnline database, which is custom designed for the BBB (Assettrac-http://www.assettrac.co.uk/).

We have nearly reached our 2500 recruitment target (2448 mothers at the end of February 2014) with nearly 1,500 trios completed and available to researchers making it a unique resource as it stands today. We have >280 complete trios from recurrent miscarriage, preterm and pre-eclampsia and >200 from FGR. We will finish recruitment and collection by June 2014 and at that time will be able to publish a detailed breakdown of the BBB cohort. To date, only six participants from the 2500 who have participated have withdrawn indicating committed interest from the participants and public and satisfaction with BBB protocol, staff and information sheets.

The BBB is a unique collection of biological samples and medical data available for national and international researchers interested in understanding pregnancy complications. We opened on the 1st November 2013 and have approved the first three applications from interested researchers and have already received preliminary enquiries from a further ten. More information on the BBB is available at http://www.ucl.ac.uk/babybiobank including the BBB protocol and biannual newsletters. We welcome your applications.

Acknowledgement

The BBB would like to thank all participating hospital staff and families.

Contribution to Authorship

Sayeda Abu-Amero is the BBB Manager and wrote the manuscript. Anna Thomas and Shawnelle White are the BBB Sample Management team, Katherine Rogers, Ana Maria Perez Miranda and Nita Solanky are the BBB Recruiting Team. Lesley Regan and Gudrun Moore are the BBB Directors. Catherine Williamson and Mark Johnson are the BBB Clinical Investigators of the participating hospital sites. All other authors commented on the manuscript.

Details of Ethical Approval

The Baby Bio Bank has Tissue Bank approval (09/H0405/30) from Trent Research Advisory Committee (27th July 2009-2014).

Funding

The BBB is funded by the Wellbeing of Women (WoW-http://www.wellbeingofwomen.org.uk/research/baby-bio-bank/), a charity dedicated to improving the health of women and babies. The funds for the BBB were raised through the Lord Mayor's Appeal in 2007-2008 and has been supplemented with generous donations from the Mothercare Foundation. The funds raised cover the cost of six full time staff to recruit the 2500 trios and prepare the sample collection over five years, as well as the equipment for preparation and storage, and the processes to properly manage the bank.

References

1. Rai R, Regan L (2006) Recurrent miscarriage. Lancet 368: 601-611.
2. Plunkett J, Muglia LJ (2008) Genetic contributions to preterm birth: implications from epidemiological and genetic association studies. Ann Med 40: 167-195.
3. Al Qahtani N (2011) Doppler ultrasound in the assessment of suspected intra-uterine growth restriction. Ann Afr Med 10: 266-271.
4. Cantwell R, Clutton-Brock T, Cooper G, Dawson A, Drife J et al. (2011) Saving Mothers' Lives: reviewing maternal deaths to make motherhood

safer:2006-2008 The Eighth Report of the Confidential Enquiries into Maternal Deaths in the UK. BJOG 118 : 1-203.

5. Nunes AP, Oliveira IO, Santos BR, Millech C, Silva LP, et al. (2012) Quality of DNA extracted from saliva samples collected with the Oragene DNA self-collection kit. BMC Med Res Methodol 12: 65.

6. Gottweis H, Gaskell G, Starkbaum J (2011) Connecting the public with biobank research: reciprocity matters. Nat Rev Genet 12: 738-739.

OPEN ACCESS Freely available online

PLOS ONE

Paternally Expressed, Imprinted Insulin-Like Growth Factor-2 in Chorionic Villi Correlates Significantly with Birth Weight

Charalambos Demetriou[1,2], Sayeda Abu-Amero[1], Anna C. Thomas[1], Miho Ishida[1], Reena Aggarwal[3], Lara Al-Olabi[1], Lydia J. Leon[1], Jaime L. Stafford[1], Argyro Syngelaki[4], Donald Peebles[3], Kypros H. Nicolaides[4], Lesley Regan[2], Philip Stanier[1], Gudrun E. Moore[1]*

1 Fetal Development and Growth Research Group, Clinical and Molecular Genetics Unit, Institute of Child Health, University College London, London, United Kingdom, **2** Department of Obstetrics and Gynaecology, St. Mary's Campus, Imperial College London, London, United Kingdom, **3** Institute for Women's Health, University College London, London, United Kingdom, **4** Harris Birthright Research Centre for Fetal Medicine, King's College Hospital, London, United Kingdom

Abstract

Context: Fetal growth involves highly complex molecular pathways. IGF2 is a key paternally expressed growth hormone that is critical for *in utero* growth in mice. Its role in human fetal growth has remained ambiguous, as it has only been studied in term tissues. Conversely the maternally expressed growth suppressor, *PHLDA*2, has a significant negative correlation between its term placental expression and birth weight.

Objective: The aim of this study is to address the role in early gestation of expression of *IGF1*, *IGF2*, their receptors *IGF1R* and *IGF2R*, and *PHLDA*2 on term birth weight.

Design: Real-time quantitative PCR was used to investigate mRNA expression of *IGF1*, *IGF2*, *IGF1R*, *IGF2R* and *PHLDA2* in chorionic villus samples (CVS) (n = 260) collected at 11–13 weeks' gestation. Expression was correlated with term birth weight using statistical package R including correction for several confounding factors.

Results: Transcript levels of *IGF2* and *IGF2R* revealed a significant positive correlation with birth weight (0.009 and 0.04, respectively). No effect was observed for *IGF1*, *IGF1R* or *PHLDA2* and birth weight. Critically, small for gestational age (SGA) neonates had significantly lower *IGF2* levels than appropriate for gestational age neonates ($p = 3{\cdot}6 \times 10^{-7}$).

Interpretation: Our findings show that *IGF2* mRNA levels at 12 weeks gestation could provide a useful predictor of future fetal growth to term, potentially predicting SGA babies. SGA babies are known to be at a higher risk for type 2 diabetes. This research reveals an imprinted, parentally driven rheostat for *in utero* growth.

Citation: Demetriou C, Abu-Amero S, Thomas AC, Ishida M, Aggarwal R, et al. (2014) Paternally Expressed, Imprinted Insulin-Like Growth Factor-2 in Chorionic Villi Correlates Significantly with Birth Weight. PLoS ONE 9(1): e85454. doi:10.1371/journal.pone.0085454

Editor: Cees Oudejans, VU University Medical Center, Netherlands

Received November 1, 2013; **Accepted** December 4, 2013; **Published** January 15, 2014

Funding: CD is funded by Save the Baby Unit (http://www.savethebabyunit.org/). PS is funded by Great Ormond Street Hospital Children's Charity (http://www.gosh.org/gen/). GEM Fetal growth and development research team is funded by the MRC (http://www.mrc.ac.uk/index.htm), Wellbeing of Women (http://www.wellbeingofwomen.org.uk/), Sparks (http://www.sparks.org.uk/) and the Wellcome Trust (http://www.wellcome.ac.uk/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gudrun.moore@ucl.ac.uk

Introduction

Fetal growth involves a complex interaction between genes and the environment, with such significant complexity that many of the molecular pathways remain to be elucidated. Normal fetal growth depends on the successful nutrient exchange between the mother and the fetus via the placenta. When this critical balance is impaired it can result in a small for gestational age (SGA) baby [1]. SGA neonates have a reported incidence of ~6% of pregnancies in developed countries and as high as 40% in some developing countries [2,3]. SGA is also associated with an increased risk of neonatal death. Although the survivors do exhibit catch up growth, it is important to note that they remain at a higher risk for common, late onset chronic diseases such as type 2 diabetes (T2D) and cardiovascular disease [4]. Hales et al., (1991) reported a graded inverse association between birth weight and T2D risk with the highest risks of T2D occurring at the lowest levels of birth weight [5]. In a more recent meta-analysis, an increased risk for T2D (OR = 1.47) is seen for SGA versus appropriate for gestational age (AGA) neonates [6].

One group of growth genes of particular interest are the imprinted genes that are found almost exclusively in Eutherian (placental) mammals. Genomic imprinting defines allele-specific, differential expression of a gene, according to its parent of origin. If the paternal allele is expressed, the maternal allele is imprinted (silenced) and vice versa. The "parental conflict hypothesis" is still the most widely accepted explanation for the evolution of genomic

imprinting [7]. It has been suggested that expression of the father's genes enhance fetal growth improving the success of the paternal genome to be passed on. In contrast, the mother's genome limits fetal growth, distributing equal resources to each of her offspring, whilst ensuring her own survival post birth allowing her to reproduce again. Studies of imprinted genes generally support this model, with one of the most striking examples being the reciprocal imprinting effects and associated growth patterns for mouse insulin-like growth factor 2 (*Igf2*) and its chelating receptor *Igf2r* [8]. In transgenic mice, loss of function of the paternally-expressed *Igf2* results in a 40% reduction in birth weight which contrasts with loss of the maternally-expressed *Igf2r*, resulting in a 30% increase in birth weight [9,10]. In addition, *IGF2* has been implicated in two imprinted human growth disorders, the overgrowth, Beckwith-Wiedemann syndrome (BWS) [11] and the growth restricting Silver-Russell syndrome (SRS) [12].

Pleckstrin homology-like domain family A member 2 (*PHLDA2*) is maternally expressed, paternally imprinted, in both humans and mice [13]. Involvement of *Phlda2* in growth and development of the placenta was demonstrated by knockout mouse models that were associated with a significant increase in placental size during mid to late gestation [14]. Studies that have investigated the human placental expression of *PHLDA2* report increased expression in SGA pregnancies and a negative association with birth weight [15,16,17].

Insulin-like growth factor 1 (*IGF1*) and its primary binding receptor *IGF1R* are not imprinted. *IGF1* exerts its growth properties on almost every cell in the body. A homozygous partial deletion of *Igf1r* in mice stunted height and weight, as well as disrupted the pubertal growth spurt. A complete inactivation of *Igf1r* is lethal in the neonatal period [18]. In contrast to *Igf2* deficient mice, restriction of growth was seen to continue into the postnatal period in the *Igf1* mutants [19].

The analyses of human placental expression of *IGF1*, *IGF2*, *IG1R* and *IGF2R* have previously been confined to samples obtained at the time of birth. *IGF1* under-expression was observed in term placental samples from SGA pregnancies [20], while mutations in the *IGF1R* gene led to abnormalities in the function of IGF1 receptors that may also slow down intrauterine and subsequent growth in humans [21].

Studies on *IGF2* have reported conflicting results. Some demonstrate no correlation between *IGF2* expression and birth weight [16] whereas others have variably shown that in SGA pregnancies compared to controls, *IGF2* expression is either increased [22], decreased [15,23,24], or similar [25], including at the protein level [26]. Moreover, studies have either shown no significant relationship between IGF2 cord serum levels and size at birth [27], or a positive effect on birth weight [28]. In others, IGF2 cord blood levels were significantly correlated with birth weight only when its interaction with IGF2R was taken into account [29].

The profound role of the IGF1 and IGF2 pathways in the regulation of fetal growth have been established largely based on experiments using the mouse as a model [9,10,18]. Analysis in humans has been hampered by the lack of available tissue to study during the course of pregnancy and has therefore relied on the use of term placenta, at a time when this tissue has become redundant and therefore the levels of gene expression may no longer reliably reflect the needs of the growing baby. To fully assess the role of these growth factors and their receptors, we have focused on an earlier developmental time point when these genes are much more likely to measure functionally relevant expression levels.

Methods

Study population

Chorionic villus sampling (CVS) was performed at 11–13 weeks of gestation in 260 singleton pregnancies that subsequently resulted in normal live births at term. The samples were collected from women undergoing CVS for prenatal diagnosis of chromosomal defects at King's College Hospital London. The excess tissue samples used for this study were obtained from women agreeing to participate in research, which was approved by the King's College Hospital Ethics Committee.

Demographic characteristics were recorded including maternal age, racial origin, smoking status, parity and body mass index as well as pregnancy outcomes such as gestational age at delivery, sex and birth weight. Other birth parameters such as placental weight were not available, nor were blood or tissue samples from the newborn baby. The pregnancies were subdivided according to the birth weight of neonates into small for gestational age (SGA) with birth weight $<10^{th}$ percentile, large (LGA) with birth weight $>90^{th}$ percentile and appropriate (AGA).

Preparation of DNA and RNA from chorionic villus samples

DNA and RNA were extracted using the iPrep Purification Instrument (Invitrogen), either by use of the iPrep^{TM} ChargeSwitch® gDNA Tissue Kit, or iPrep^{TM} PureLink^{TM} Total RNA and Trizol® Plus RNA kit including DNAse treatment, according to the manufacturer's instructions.

Reverse transcription

Reverse transcriptase (RT) methodology was based on a standard protocol using M-MLV reverse transcriptase and random primer hexamers (Promega). Primers for the housekeeping gene *β-actin* (*ACTB*) were used to check the integrity of the cDNA and ensure no DNA contamination. The forward (F) and reverse (R) primers spanned an intron and the sequences are: *β*-actin.F- gtcttcccctccatcgtg and *β*-actin.R- ggtcatcttctcgcggttg.

Polymerase Chain Reaction (PCR)

Genomic DNA and cDNA from CVS were amplified by PCR before sequencing. Primers (5′ to 3′ sequence) used for genomic DNA are IGF2.F- aacaccccacaaaagctcag; IGF2.R- tgcatggattttggttttca; IGF2R.F- gaaacacaaaacctacgacc; IGF2R.R- agaacccaaaagagccaacc; PHLDA2.F- caaaccccgcacgccatgag and PHLDA2.R- ctgtgcccattgcaaataaatc. The same primers were used for cDNA with the exception of IGF2R.R- cctttggagtacgtgacaac. 20 ul reactions were set up and thermal cycling conditions were 94°C for 5 min, 94°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec for 35 cycles, and 72°C for 2 min.

Imprinting analysis

Imprinting analysis of *IGF2*, *IGF2R* and *PHLDA2* in CVS was carried out by DNA sequence analysis of expressed single nucleotide polymorphisms (SNP) in CVS gDNA after amplification using specific primers (IGF2.F- aacaccccacaaaagctcag; IGF2R.R- cctttggagtacgtgacaac and PHLDA2.F- caaaccccgcacgccatgag). Corresponding cDNA samples were screened in patients heterozygous for the *IGF2* A/G (rs680), *IGF2R* A/G (rs1805075) and *PHLDA2* A/G (rs1056819) SNPs (http://www.ncbi.nlm.nih.gov/projects /SNP/). Sequencing reactions were prepared according to the manufacturer's instructions (Applied Biosystems) using the ABI Prism Big Dye terminator cycle sequencing ready reaction kit (BDT v1.1). Sequencing products were run on an ABI

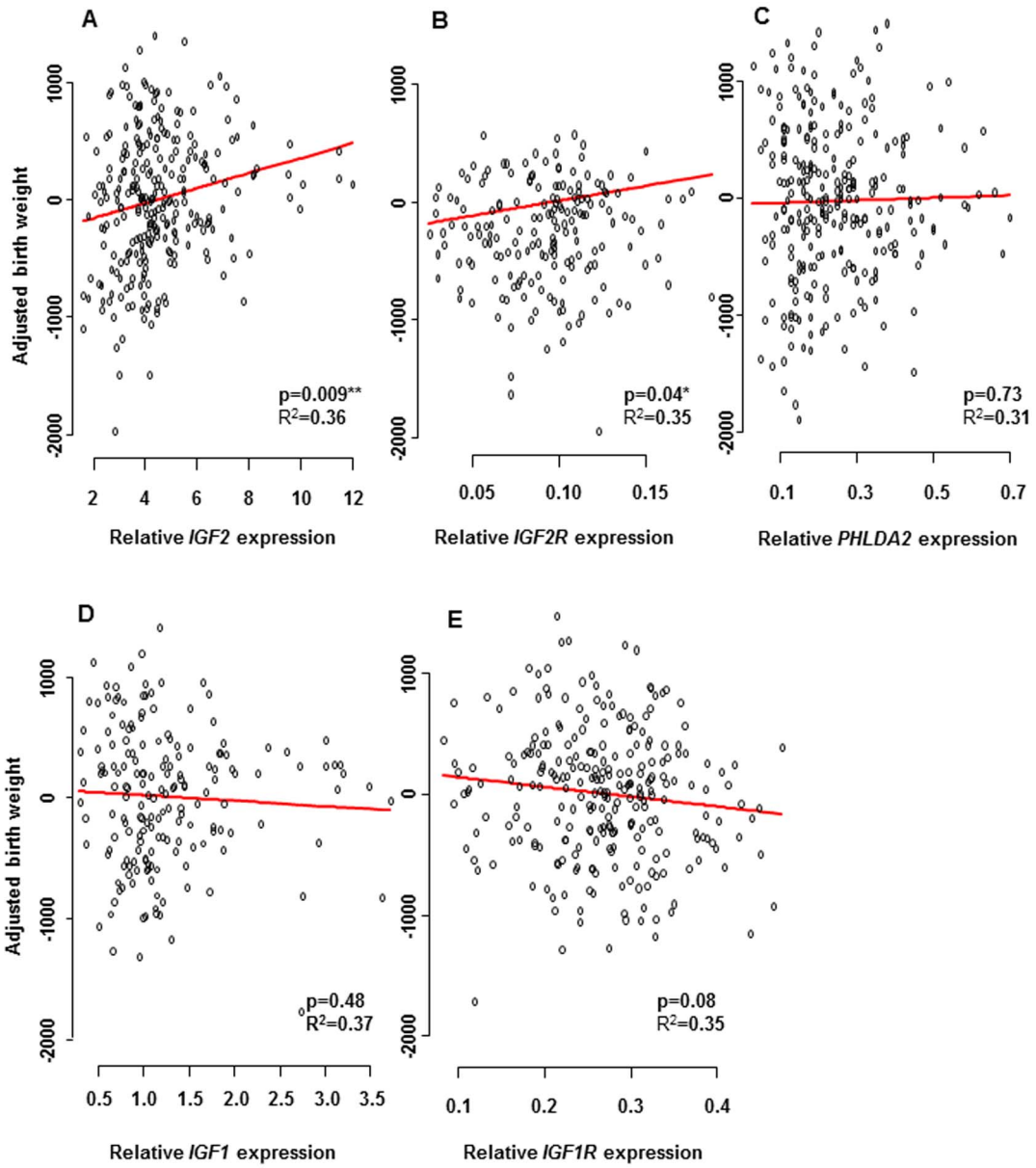


Figure 1. mRNA expression levels of *IGF2, IGF2R, PHLDA2, IGF1* and *IGF1R* in chorionic villi. The expression levels of *IGF2*, *IGF2R*, *PHLDA2*, *IGF1* and *IGF1R* after standardization to the endogenous control gene *L19*, in relation to birth weight corrected for parity, sex, GA at term, maternal BMI and smoking status. Significant associations were observed for CVS expression of A) *IGF2* (p = 0.009) and B) *IGF2R* (p = 0.04). No significant association was found for C) *PHLDA2* (p = 0.73), for D) *IGF1* (p = 0.48) or for E) *IGF1R* (p = 0.08).
doi:10.1371/journal.pone.0085454.g001

Prism 3730 DNA analyzer, and the read-out was analysed with Sequencher™ v4.8 (Gene Codes Corporation).

Real-time quantitative PCR

The quantitative expression analysis of the genes of interest, *IGF1*, *IGF2*, *IGF1R*, *IGF2R* and *PHLDA2* as well as the endogenous control gene *L19* (a housekeeping gene ubiquitously expressed in the placenta) [16] was determined by real-time quantitative PCR (RTqPCR) with SYBR Green (ABI) using the StepOnePlus Real-Time PCR System (ABI). Primers (5′ to 3′ sequence) used for quantitative analysis are IGF1.F- ggaggctggagatgtattgc; IGF1.R- acttgcttctgtcccctcct; IGF2.F- cgagagggacgtgtcgacc; IGF2.R- ggactgcttccaggtgtcata; IGF2R.F- ccggcgtgctctgga; IGF2R.R- ccagagggtcacagtggaaga; IGF1R.F- ccaagggtgtggtgaaagat; IGF1R.R- tccatgatgaccagtgttgg; PHLDA2.F- ccatccccgcagccccaaacc; PHLDA2.R- ccacgtcctagcctgggtcc; L19.F- gcggaagggtacagccaat and L19.R- caggctgtgatacatgtggcg. All primer sets were free of primer-dimer products. The RTqPCR assays were run in triplicate for each sample, with the gene of interest and housekeeping gene run on the same 96-well plates. A control pool of CVS cDNA was also included on each plate. Amplification conditions include initial incubation at 95°C for 10 min and repetitive denaturation at 95°C for 15 sec and annealing at 60°C for 1 min for 40 cycles.

After amplification, quantitative expression levels were obtained using the StepOne software (version 2.1). All triplicate cycle threshold (C_T) values were within 1 C_T of each other. The quantitative values for each triplicate were averaged and the

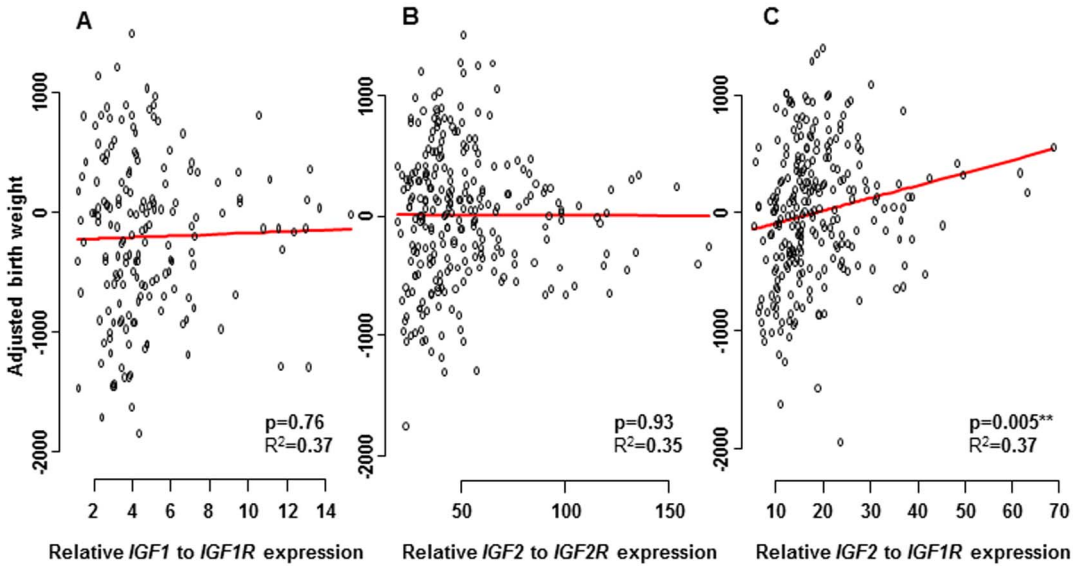


Figure 2. mRNA expression levels of *IGF1/IGF1R*, *IGF2*/*IGF2R* and *IGF2/IGF1R* in chorionic villi. The expression levels of *IGF1/IGF1R*, *IGF2/IGF2R* and *IGF2/IGF1R* after standardization to the endogenous control gene *L19*, in relation to birth weight corrected for parity, sex, GA at term, maternal BMI and smoking status. No significant association was found for A) the ratio of *IGF1* to *IGF1R* (p = 0.76), or for B) the ratio of *IGF2* to *IGF2R* (p = 0.93). Significant association was observed for CVS expression of C) the ratio of *IGF2* to *IGF1R* (p = 0.005) and birth weight.
doi:10.1371/journal.pone.0085454.g002

relative expression of the genes of interest was determined by a ratio of their expression to that of the *L19* housekeeping gene for the same sample.

Statistical analysis

Standard multiple linear regression models were used to examine the correlation of *IGF1*, *IGF2*, *IGF2R*, *IGF1R* and *PHLDA2* relative expression values to birth weight after adjustment for gestational age at delivery, sex, parity, maternal age, smoking status and maternal body mass index. Any outliers more than 2 standard deviations from the mean were removed. The analysis of variance (one-way ANOVA) was used to compare the gene expression means between the three different categorical birth weight groups (SGA, AGA, LGA) taking into consideration any confounding factors. Statistical package "R" (version 2.13.0) was used for the analyses and for calculating adjusted R-squared (R^2) values. Significance was defined as $p<0.05$.

Results

In this study, we have investigated the relationship between *IGF1*, *IGF2*, *IGF2R*, *IGF1R* and *PHLDA2* mRNA levels in first-trimester placental tissue with the birth weight of the resultant term new born babies. We correlate this data taking into account the carefully recorded birth parameters that may have a confounding effect (gestational age, sex, parity, maternal BMI and smoking status; figure S1). Regression analysis was performed after standardization to the endogenous control gene *L19*, in relation to birth weight corrected for confounding factors.

Significant associations were observed for expression of *IGF2* (p = 0.009) and *IGF2R* (p = 0.04) between CVS tissue and birth weight (figure 1A–B). Importantly the range of *IGF2* in relative expression units was from 1.5–12 with one relative expression unit being equivalent to a change in birth weight of 63 g. No association was found for *PHLDA2* (p = 0.73), *IGF1* (p = 0.48) or *IGF1R* (p = 0.08) between CVS tissue and birth weight (figure 1C–E).

As both IGF1 and IGF2 bind to the IGF1R and only IGF2 binds to the IGF2R we decided to look at this relationship between the ligands and their receptors and any correlation they might have with birth weight. The ratio of *IGF2 to IGF1R* (p = 0.005) was significantly associated with birth weight but no association was found for the ratio of *IGF1 to IGF1R* (p = 0.76) or the ratio of *IGF2 to IGF2R* (p = 0.93) between CVS tissue and birth weight (figure 2).

To investigate the whole growth spectrum we subdivided the pregnancies into small for gestational age (SGA; n = 50) with birth weight $<10^{th}$ percentile, large (LGA; n = 65) with birth weight $>90^{th}$ percentile and appropriate (AGA; n = 145). In the SGA group compared to the AGA group, *IGF2* expression was reduced by statistically significant amounts ($p = 3.6\times10^{-7}$), despite no significant differences in the expression levels of individual receptors (*IGF2R*, p = 0.76 or *IGF1R*, p = 0.15). In the LGA group *IGF2R* expression was found to be higher than in the AGA group (p = 0.02) and *IGF1R* expression was lower than in the SGA group (p = 0.05) (figure 3).

Given that *IGF2* and *PHLDA2* are both known to be imprinted in mouse and human [8,13,30], we wanted to confirm that monoallelic expression was maintained in CVS material, since reversion to biallelic expression may impact on mRNA levels. Furthermore, *IGF2R* in the human population is reported to be monoallelically expressed (i.e. imprinted) in only 10% of individuals, which is in contrast to the mouse, where it is fully imprinted [31]. We therefore wished to confirm its imprinting status in CVS material and to investigate for any potential skewing. The transcribed SNPs rs680 (*IGF2*), rs1805075 (*IGF2R*) and rs1056819 (*PHLDA2*) were used to report on allele-specific expression and thus determine imprinting status of *IGF2*, *IGF2R* and *PHLDA2*. Genomic DNA extracted from the CVS from 200 patients were tested for heterozygosity at these SNPs, with 40 samples found to be informative at rs680 for *IGF2*, 24 samples at rs1805075 for *IGF2R* and 21 samples at rs1056819 for *PHLDA2*. The 40 informative *IGF2* individuals and the 21 informative *PHLDA2* individuals were all found to be monoallelically expressed. For the 24 samples informative for *IGF2R*, 21 (88%)

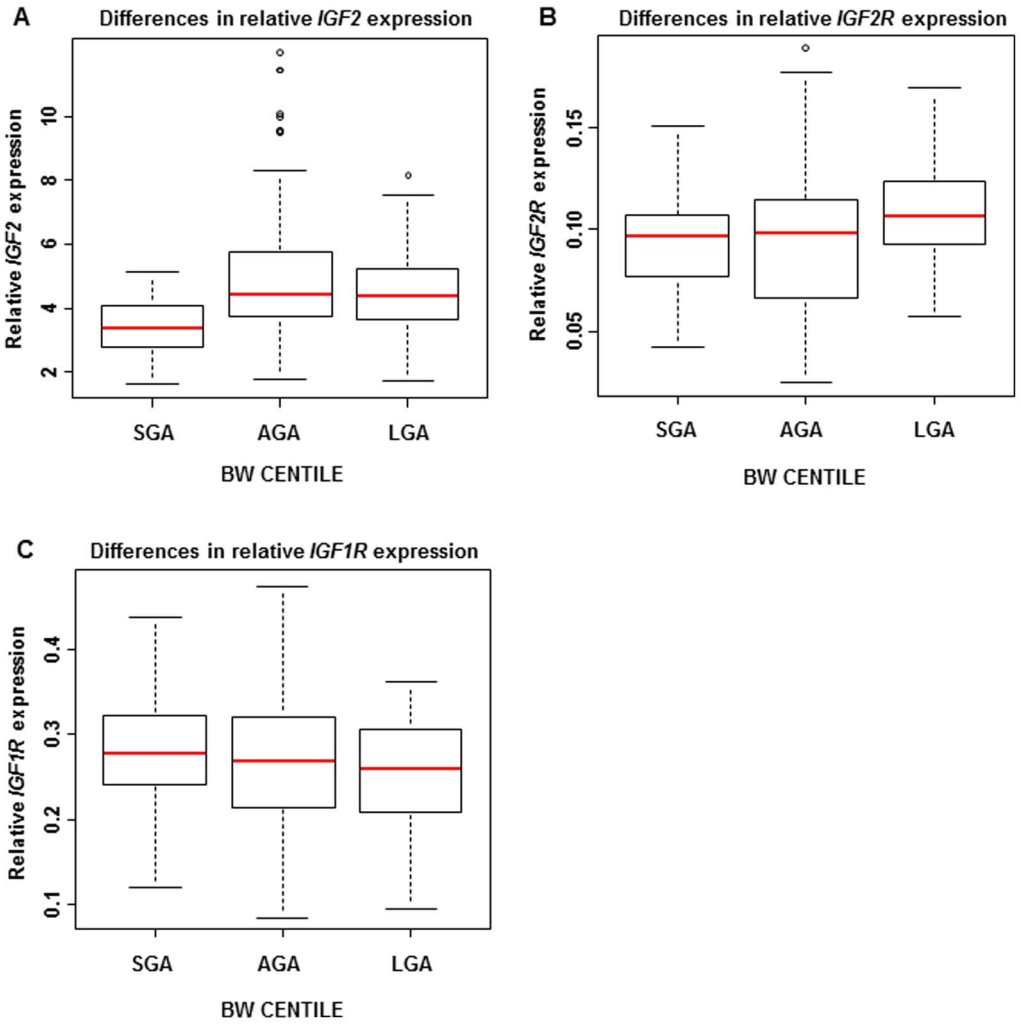


Figure 3. mRNA expression levels of *IGF2*, *IGF2R*, and *IGF1R* in chorionic villi according to BW centile. Relative *IGF2*, *IGF2R*, and *IGF1R* expression levels in the pregnancies with small (SGA), appropriate (AGA) and large (LGA) neonates. In the SGA group, compared to the AGA group, A) *IGF2* expression was lower ($p = 3{\cdot}6 \times 10^{-7}$), but there was no significant difference in the expression levels of B) *IGF2R* ($p = 0{\cdot}76$) or C) *IGF1R* ($p = 0{\cdot}15$). In the LGA group B) *IGF2R* expression was higher than in the AGA group ($p = 0{\cdot}02$) and C) *IGF1R* expression was lower than in the SGA group ($p = 0{\cdot}05$).
doi:10.1371/journal.pone.0085454.g003

were biallellically expressed for *IGF2R* and 3 (12%) were monoallelic, which is similar to the expected population frequency (figure 4). Thus, skewed monoallelic/biallelic expression was unlikely to be a factor reflected in the relative expression levels associated with birth weight.

Discussion

Previous studies have linked *IGF2* expression to birth weight, by showing that *IGF2* in term placenta is decreased in SGA pregnancies compared to controls [23,24]. However, the first trimester *IGF2* expression data reported in this study, provides the first evidence for the role of this paternally expressed imprinted gene as an *in utero* fetal growth enhancer this early in human pregnancy. This is compatible with results previously described in animal studies [9,32]. For example, in mouse experiments, paternally expressed *Igf2* was reported to control 40% of fetal growth, with the maternally expressed gene *Igf2r* limiting fetal growth [10,33]. In humans, *IGF2* is consistently maternally imprinted and therefore a paternal-expression driven growth promoter. However, in humans, *IGF2R* is polymorphic for its imprinting status with 90% of individuals showing biallelic expression [31]. This indicates a possible shift in the mechanism of its regulation of expression away from that used in the mouse. Interestingly, here, the samples showing monallelic expression of *IGF2R* had similar levels to those showing biallelic expression.

As fetal size is subject to both positive and negative regulation, *IGF2* and *IGF2R* genes can exert opposite forces on the fetus, achieving a fine degree of control over the growth process. As levels of *IGF2* increase in the fetus, the levels of *IGF2R* can also increase, resulting in clearance of IGF2 from plasma and tissue fluids, correcting and maintaining levels of IGF2 in the circulation. This acts as a fine-tuning rheostat designed to prevent babies from overgrowth in order to maintain a normal growth trajectory. This is also supported by the fact that no association was observed between the *IGF2*/*IGF2R* expression levels and birth weight, as the ligand and the receptor work tightly together to regulate growth. The same applies for IGF1 and its regulating receptor IGF1R.

In this study, expression of *IGF1* in chorionic villi was not associated with birth weight outcome. These results directly

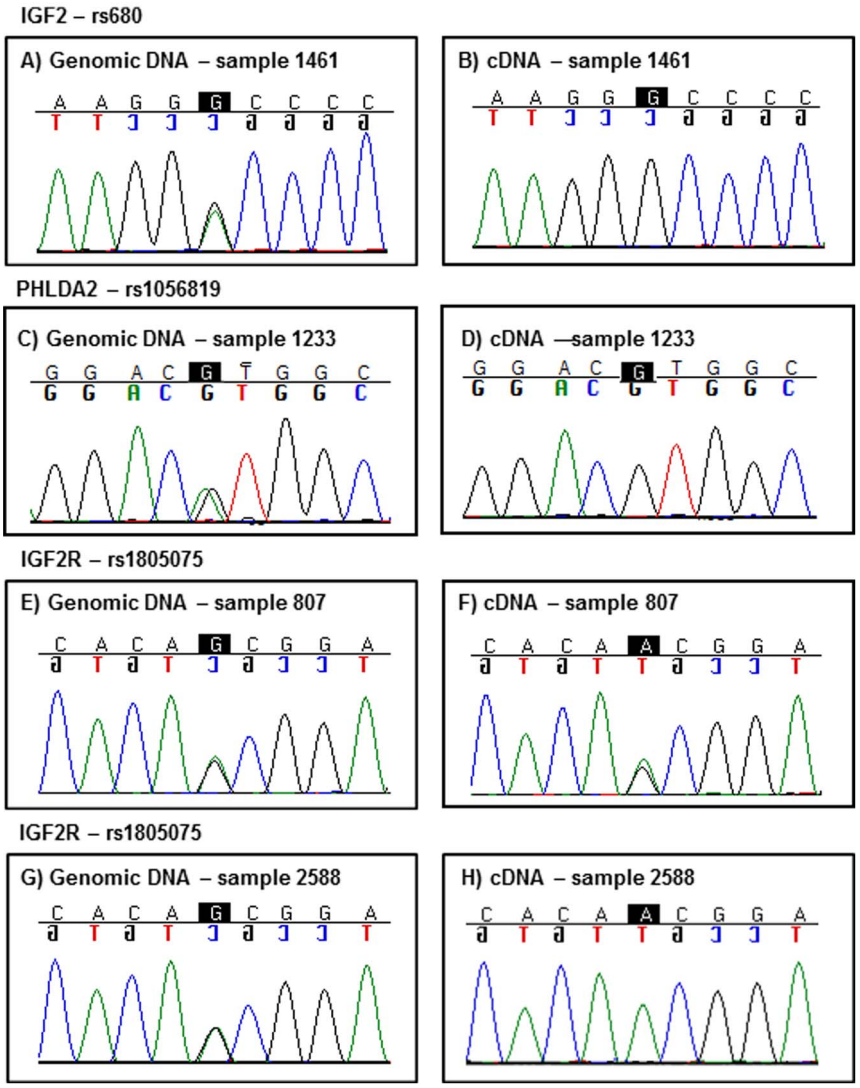


Figure 4. Imprinting analysis of *IGF2, PHLDA2* and *IGF2R*. Representative sequencing chromatograms are shown from one of the 40 informative CVS samples for the *IGF2* A/G polymorphism (rs680) (A, B), one (from 21) for the *PHLDA2* A/G polymorphism (rs1056819) (C, D) and two (from 24) for the *IGF2R* A/G polymorphism (rs1805075) (E, F, G, H). Heterozygous genomic CVS DNA sequence is shown in A), C), E) and G). While monoallelic expression from the corresponding cDNA samples are shown in B), D) and H). Monoallelic expression in cDNA was found in all informative *IGF2* and *PHLDA2* samples as well as 12% of *IGF2R* samples. Biallelic expression, as shown in F), was found in the majority (88%) of informative cDNA samples for *IGF2R*.
doi:10.1371/journal.pone.0085454.g004

contradict findings from studies involving mouse models which highlighted *IGF1* role in the regulation of both pre- and postnatal growth [34]. There also remains a strong possibility that *IGF1* is indeed relevant to intrauterine growth in the later stages of pregnancy as it is preparing the baby for a postnatal life. While fetal chorionic fetal samples in the present cohort proved an opportunity to assess correlation between gene expression and fetal growth in early gestation (11–13 weeks), this narrow window might not be the best to investigate *IGF1* expression, as IGF1 does not appear in the fetal circulation until later [35].

Although previous studies have shown that placental *PHLDA2* expression is negatively associated with birth weight [16,17], in our study no statistically significant association was observed between *PHLDA2* expression levels in CVS and birth weight. This suggests that maternally expressed *PHLDA2* is suppressing the baby's growth later in pregnancy rather than early on. Previous studies investigating term placental *IGF2* expression have shown inconclusive results. However, in our study a statistically significant association was observed between *IGF2* expression levels in CVS and birth weight, suggesting that the paternal genome is promoting the baby's growth earlier in pregnancy. Therefore, the two parental genomes appear to be acting at different times during pregnancy to control the fetal weight. This supports the idea that whilst increased fetal growth is important early on, it must still require careful regulation by the mother to ensure a successful birth.

In our study, SGA neonates had significantly lower *IGF2* expression levels compared to AGA neonates, but no differences were observed between the levels of the receptors *IGF2R* and *IGF1R*. This highlights the importance of the *IGF2* ligand rather than the receptor level in determining size of the neonate in SGA pregnancies. This is in agreement with previous studies investigating *IGF2* mRNA levels in the placentas from growth-restricted pregnancies [23,24].

To better understand the causes and consequences of fetal growth restriction as a human pregnancy complication, we focused on available tissue from the first trimester, as this time window is more likely to accurately reflect the *in utero* growth potential. We used a larger sample size and we investigated the whole growth spectrum including not only SGA but also LGA neonates. We reasoned that extending the investigation of these genes to samples displaying macrosomic birth weight would also reveal further correlations between growth parameters and gene expression. In this study, LGA neonates had higher *IGF2R* expression levels compared to AGA neonates and lower *IGF1R* levels compared to SGA babies. Similar results have been reported in placenta tissue for *IGF1R* mRNA levels alone [36], and this suggests that fetuses with high birth weight are producing more *IGF2R* and correspondingly less *IGF1R*, which could remove *IGF2*. This balance may represent another important compensatory mechanism in response to fetal overgrowth. This is also supported by the fact that a significant association was observed between the ratio of *IGF2* to *IGF1R* and birth weight, suggesting that the levels of IGF1R in LGA neonates decrease as their IGF2 levels increase, to avoid any further increase in size.

The "small baby syndrome hypothesis" proposed by Barker et al., (1993) suggests that there is an inverse linear relation between birth weight and T2D [37]. This is also supported by a large meta-analysis [38]. The mechanisms by which birth weight is related to T2D is still under debate, Barker et al. claim that this relationship reflects long-term consequences of under-nutrition *in utero* [37], whereas others do not see under-nutrition as playing a significant role [39]. Interestingly, specific fetal *IGF2* paternal haplotypes are linked to higher maternal glucose levels that increase the risk of gestational diabetes [40]. Nevertheless the amount of IGF2 available from the placenta early in the first trimester is likely to be a major factor in promoting growth.

An important aim for antenatal care is the prediction, detection and treatment of anomalous fetal growth. We know that variation in size at birth results from interaction between fetal genetic factors and the maternal genetic and uterine environment. In this study we also implicate the important role of the father's genes in fetal growth *in utero*. Understanding the developmental role, function and parent-of-origin effect of critical imprinted genes during human fetal growth will make an important contribution to effective clinical evaluation of growth disorders and their association with longer term, metabolically related, health risks such as T2D. Determination of paternal *IGF2* expression in the first trimester, potentially in maternal blood, may act as a predictor of fetal growth trajectory for later in the pregnancy. An early growth biomarker could be invaluable to alert Obstetricians and Neonatologists towards closer 'at risk' pregnancy surveillance.

Supporting Information

Figure S1 Confounding factors. Correlation of birth weight with gestational age at term ($p = 4.8\times10^{-15}$; Fig. S1A), with maternal BMI ($p = 0.0012$; Fig. S1B), with gender ($p = 1.4\times10^{-7}$; Fig. S1C), with parity ($p = 7.5\times10^{-5}$; Fig. S1D), and maternal smoking status ($p = 0.33$; Fig S1E).
(TIF)

Author Contributions

Conceived and designed the experiments: CD SAA ACT PS GEM. Performed the experiments: CD SAA ACT MI RA LA LJL JLS. Analyzed the data: CD MI RA LA LJL JLS. Contributed reagents/materials/analysis tools: AS DP KHN LR PS GEM. Wrote the paper: CD SAA ACT MI RA LA LJL JLS AS DP KHN LR PS GEM. Provided chorionic villus samples: KHN. Gathered the sample's clinical data: AS. Provided obstetric advice and discussion: DP LR.

References

1. Pollack RN, Divon MY (1992) Intrauterine growth retardation: definition, classification, and etiology. Clin Obstet Gynecol 35: 99–107.
2. Brodsky D, Christou H (2004) Current concepts in intrauterine growth restriction. J Intensive Care Med 19: 307–319.
3. Albertsson-Wikland K, Wennergren G, Wennergren M, Vilbergsson G, Rosberg S (1993) Longitudinal follow-up of growth in children born small for gestational age. Acta Paediatr 82: 438–443.
4. Barker DJ (1992) Fetal growth and adult disease. Br J Obstet Gynaecol 99: 275–276.
5. Hales CN, Barker DJ, Clark PM, Cox LJ, Fall C, et al. (1991) Fetal and infant growth and impaired glucose tolerance at age 64. BMJ 303: 1019–1022.
6. Harder T, Rodekamp E, Schellong K, Dudenhausen JM, Plagemann A (2007) Birth weight and subsequent risk of type 2 diabetes: A meta-analysis. Am J Epidemiol 165: 849–857.
7. Moore T, Haig D (1991) Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet 7: 45–49.
8. Willison K (1991) Opposite imprinting of the mouse Igf2 and Igf2r genes. Trends Genet 7: 107–109.
9. DeChiara TM, Efstratiadis A, Robertson EJ (1990) A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature 345: 78–80.
10. Wang ZQ, Fung MR, Barlow DP, Wagner EF (1994) Regulation of embryonic growth and lysosomal targeting by the imprinted Igf2/Mpr gene. Nature 372: 464–467.
11. Weksberg R, Shen DR, Fei YL, Song QL, Squire J (1993) Disruption of insulin-like growth factor-2 in Beckwith Wiedemann syndrome. Nat Genet 5:143–149.
12. Eggermann T, Meyer E, Obermann C, Heil I, Schuler H, et al. (2005) Is maternal duplication of 11p15 associated with Silver-Russell syndrome? J Med Genet 42: e26.
13. Qian N, Frank D, O'Keefe D, Dao D, Zhao L, et al. (1997) The IPL gene on chromosome 11p15.5 is imprinted in humans and mice and is similar to TDAG51, implicated in Fas expression and apoptosis. Hum Mol Genet 6: 2021–2029.
14. Frank D, Fortino W, Clark L, Musalo R, Wang W, et al. (2002) Placental overgrowth in mice lacking the imprinted gene Ipl. Proc Natl Acad Sci USA 99: 7490–7495.
15. McMinn J, Wei M, Schupf N, Cusmai J, Johnson EB, et al. (2006) Unbalanced placental expression of imprinted genes in human intrauterine growth restriction. Placenta 27: 540–549.
16. Apostolidou S, Abu-Amero S, O'Donoghue K, Frost J, Olafsdottir O, et al. (2007) Elevated placental expression of the imprinted PHLDA2 gene is associated with low birth weight. J Mol Med 85: 379–387.
17. Ishida M, Monk D, Duncan AJ, Abu-Amero S, Chong J, et al. (2012) Maternal inheritance of a promoter variant in the imprinted PHLDA2 gene significantly increases birth weight. Am J Hum Genet 90: 715–719.
18. Klammt J, Pfaffle R, Werner H, Kiess W (2008) IGF signaling defects as causes of growth failure and IUGR. Trends Endocrinol Metab 19; 197–205.
19. Le Roith D, Scavo L, Butler A (2001) What is the role of circulating IGF-I? Trends Endocrinol Metab 12: 48–52.
20. Koutsaki M, Sifakis S, Zaravinos A, Koutroulakis D, Koukoura O, et al. (2011) Decreased placental expression of hPGH, IGF-I and IGFBP-1 in pregnancies complicated by fetal growth restriction. Growth Horm IGF Res 21: 31–36.
21. Abuzzahab MJ, Schneider A, Goddard A, Grigorescu F, Lautier C, et al. (2003) IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. N Engl J Med 349: 2211–22.
22. Abu-Amero SN, Ali Z, Bennett P, Vaughan JI, Moore GE (1998) Expression of the insulin-like growth factors and their receptors in term placentas: a comparison between normal and IUGR births. Mol Reprod Dev 49: 229–235.
23. Koukoura O, Sifakis S, Soufla G, Zaravinos A, Apostolidou S, et al. (2011) Loss of imprinting and aberrant methylation of IGF2 in placentas from pregnancies complicated with fetal growth restriction. Int J Mol Med 28: 481–487.
24. Guo L, Choufani S, Ferreire J, Smith A, Chitayat D, et al. (2008) Altered gene expression and methylation of the human chromosome 11 imprinted region in small for gestational age (SGA) placentae. Developmental Biology 320: 79–91.
25. Street ME, Seghini P, Fieni S, Ziveri MA, Volta C, et al. (2006) Changes in interleukin-6 and IGF system and their relationships in placenta and cord blood in newborns with fetal growth restriction compared with controls. Eur J Endocrinol 155: 567–574.
26. Laviola L, Perrini S, Belsanti G, Natalicchio A, Montrone C, et al. (2005) Intrauterine growth restriction in humans is associated with abnormalities in placental insulin-like growth factor signaling. Endocrinology 146: 1498–1505.
27. Klauwer D, Blum WF, Hanitsch S, Rascher W, Lee PD, et al. (1997) IGF-I, IGF-II, free IGF-I and IGFBP-1, -2 and -3 levels in venous cord blood:

relationship to birthweight, length and gestational age in healthy newborns. Acta Paediatr 86: 826–833.
28. Smerieri A, Petraroli M, Ziveri MA, Volta C, Bernasconi S, et al. (2011) Effects of cord serum insulin, IGF-II, IGFBP-2, IL-6 and cortisol concentrations on human birth weight and length: pilot study. PLoS One 6: e29562.
29. Ong K, Kratzsch J, Kiess W, Costello M, Scott C, et al. (2000) Size at birth and cord blood levels of insulin, insulin-like growth factor I (IGF-I), IGF-II, IGF-binding protein-1 (IGFBP-1), IGFBP-3, and the soluble IGF-II/mannose-6-phosphate receptor in term human infants. The ALSPAC Study Team. Avon Longitudinal Study of Pregnancy and Childhood. J Clin Endocrinol Metab 85: 4266–4269.
30. Giannoukakis N, Deal C, Paquette J, Goodyer CG, Polychronakos C (1993) Parental genomic imprinting of the human IGF2 gene. Nat Genet 4: 98–101.
31. Monk D, Arnaud P, Apostolidou S, Hills FA, Kelsey G, et al. (2006) Limited evolutionary conservation of imprinting in the human placenta. Proc Natl Acad Sci USA 103: 6623–6628.
32. Constância M, Hemberger M, Hughes J, Dean W, Ferguson-Smith A, et al. (2002) Placental-specific IGF-II is a major modulator of placental and fetal growth. Nature 417: 945–948.
33. Lau MM, Stewart CE, Liu Z, Bhatt H, Rotwein P, et al. (1994) Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. Genes Dev 8: 2953–2963.
34. Le Bouc Y, Gircquel C, Holzenberger M (2003) Physiology of somatotropic axis: interest of gene inactivation experiments. Bull Acad Natl Med 187: 1225–43.
35. Demendi C, Börzsönyi B, Nagy ZB, Rigó J Jr, Pajor A, et al. (2011) Gene expression patterns of insulin-like growth factor 1, 2 (IGF-1, IGF-2) and insulin-like growth factor binding protein 3 (IGFBP-3) in human placenta from preterm deliveries: influence of additional factors. Eur J Obstet Gynecol 160: 40–44.
36. Iniguez G, Gonzalez CA, Argandona F, Kakarieka E, Johnson MC, et al. (2010) Expression and protein content of IGF-I and IGF-I receptor in placentas from small, adequate and large for gestational age newborns. Horm Res Paediatr 73: 320–327.
37. Barker DJ, Hales CN, Fall CH, Osmond C, Phipps K, et al. (1993) Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. Diabetologia 36: 62–67.
38. Whincup PH, Kaye SJ, Owen CG, Huxley R, Cook DG, et al. (2008) Birth weight and risk of type 2 diabetes A systematic review. JAMA 300: 2886–2897.
39. Hofman PL, Regan F, Jackson WE, Jefferies C, Knight DB, et al. (2004) Premature birth and later insulin resistance. N Engl J Med 351: 2179–2186.
40. Petry CJ, Seear RV, Wingate DL, Manico L, Acerini CL, et al. (2011) Associations between paternally transmitted fetal IGF2 variants and maternal circulating glucose concentartions in pregancy. Diabetes 60: 3090–3096.

Microbiome analysis of human gestational tissues in normal and complicated pregnancies

Lydia Leon, Ronan Doyle, Nigel Klein, Philip Stanier, Gudrun Moore

UCL

Aims

- Explore how human endogenous bacteria contribute to outcomes in pregnancy within a cohort of complicated and normal pregnancies that deliver in the UK.
- Quantify the presence/absence of total bacteria as well as relative species abundances within gestational tissues.
- Explore variation in bacterial load between control and preterm group, but also within preterm group (e.g. spontaneous vs. indicated vs. pPROM preterms).
- Investigate the relationship between bacterial abundance and diversity profiles, and the maternal immune response.

Background

Around 70,000 babies are born preterm (before 37 completed weeks gestation) in the UK annually. Numerous morbidities are associated with preterm birth in the neonatal period and beyond, and it is the leading cause of neonatal death globally.

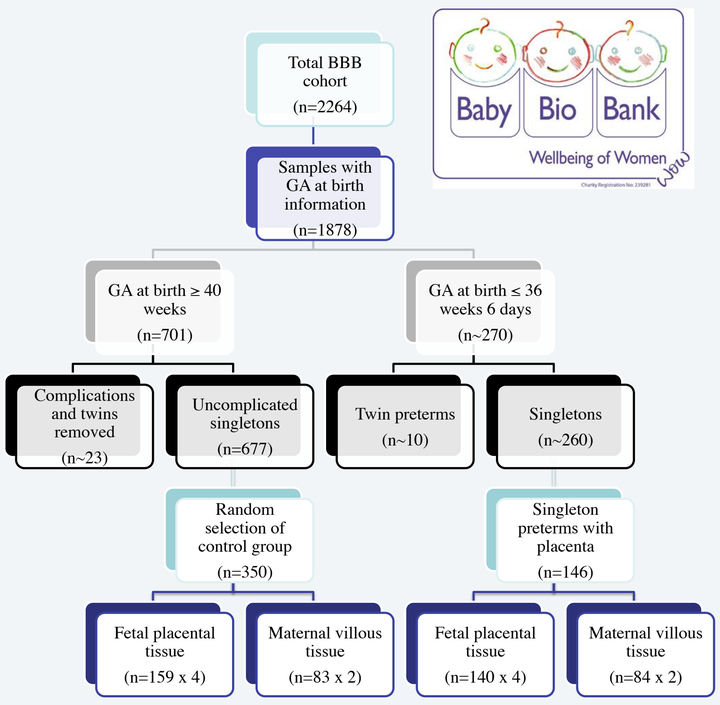
A large proportion of preterm births are characterised by signs of infection and/or inflammation in the uterus, the vagina and at the maternal-fetal interface, with increased prevalence and diversity of bacteria often observed in preterm versus term gestational tissues (e.g. fetal membrane, amniotic fluid, placenta, umbilical cord). Up-regulation of several proteins involved in inflammatory pathways have also been observed in pre-term versus term placenta e.g. IL-1β, TNFα, IL-8.

Study Population

The Baby Bio Bank (BBB):
A resource that contains clinical data and biological samples from nearly 3000 pregnancies at 3 hospitals across London.

The BBB is a cohort of mainly white Caucasian mothers experiencing both normal and complicated pregnancies (growth restriction, pre-eclampsia, multiple miscarriage, preterm birth).

For the preterm cohort, cases delivered before 37 completed weeks gestation. Controls were randomly selected from deliveries at or after 40 weeks, following an uncomplicated pregnancy. Cases can be further divided according to nature of labour (spontaneous or non-spontaneous) and membrane rupture method.



Materials and Methods

DNA extraction from tissue

- 3 x 50mg tissue excised from two points on fetal side of placenta and a mixture of sites from maternal villous tissue.
- Care taken to avoid possible events of contamination at all stages of collection and extraction.
- *Qiagen mini-prep* for blood and tissue used according to manufacturers protocol with additional bead-beating step added to ensure sufficient lysis of bacterial cell walls.

Real time PCR absolute quantification using *SYBR-Green*

- Broad range primers anneal to highly conserved regions of bacterial specific 16s rRNA gene to enable cross-species amplification of any bacterial DNA present in placental samples.
- E-coli standard curve run on each real time plate to enable relative quantification of sample bacterial population against this standard of known bacterial DNA concentration.

Preliminary Results

At present, only a third of placental samples have been analysed and the following data is therefore preliminary.

Initial real time experiments were complicated by need to limit inhibitory effect of human DNA (~99% of sample) whilst maximising sensitivity of assay to small amount of target bacterial 16s sequence (figure 1).

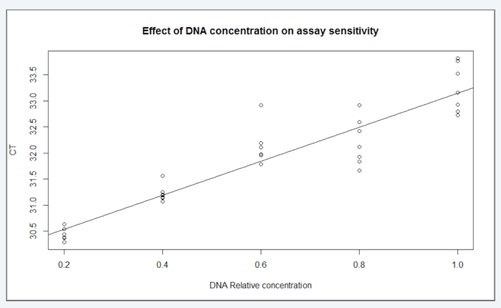


Figure 1- As concentration of starting template increases, sensitivity of assay decreases – shown by increase in average CT of same samples

Preliminary Results

This inhibitory effect of higher template concentration is observed in samples of human DNA spiked with known quantities of *E. coli* DNA but not when only *E. coli* standards are used (figure 2). As a result of this methodological work, it was decided that all subsequent work would be normalised to 25ng/ul to minimise this inhibitory effect but maximise sensitivity to any endogenous bacterial DNA in sample.

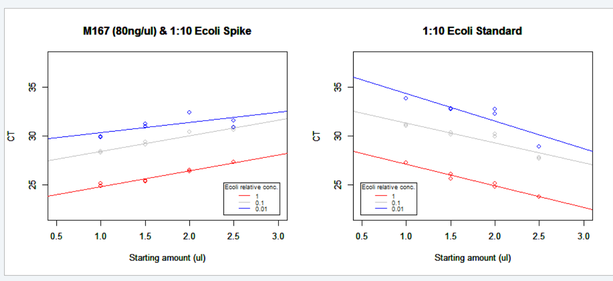


Figure 2 – Inhibitory effect of mixed sample human DNA on bacterial 16s target amplification

From figure 3 it is clear that as gestational age increases in the cohort, the proportion of vaginal births compared to caesarean sections (CS) increases. Despite concerns over variation in levels of bacterial contamination depending on mode of delivery (i.e. higher exposure in vaginal births), there is no evidence so far of an effect of delivery method on absolute bacterial prevalence in placental tissues in this cohort (figure 4).

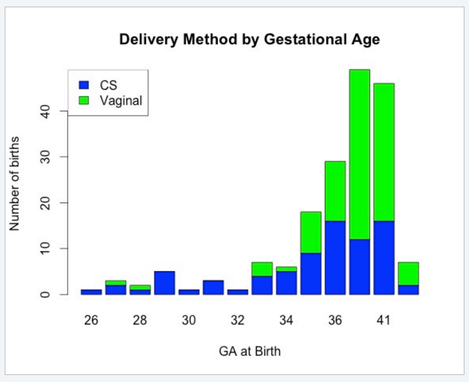


Figure 3 – Delivery Method by GA

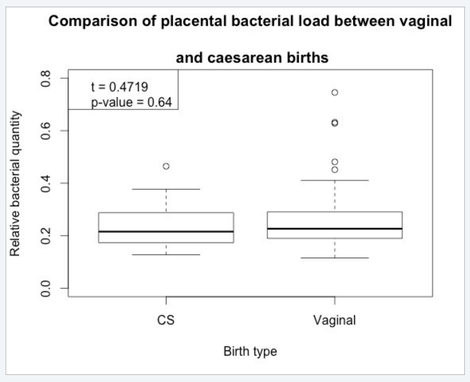


Figure 4 – Delivery method and bacterial load

At present, there is no evidence of a difference in the absolute quantity of bacterial DNA between control and preterm placentas. However, when preterms are further divided into spontaneous preterm births (sPTB) and non-spontaneous (non-sPTB), sPTB placentas appear to have a higher bacterial load than both controls and nsPTBs – the biggest difference is between the two preterm groups (figure 5). This highlights the need to consider carefully the categorization of this cohort – sPTB will likely have a different aetiology than non- sPTB. This is in line with evidence from independent studies that bacterial infection is more likely to be a causal factor in spontaneous as opposed to indicated preterm births.

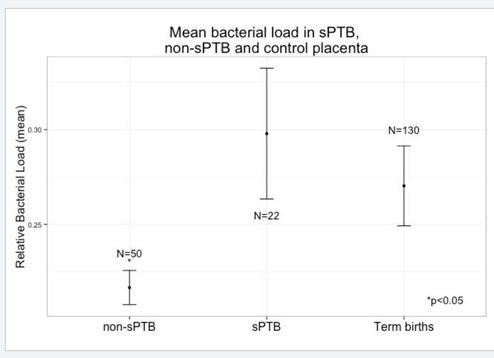


Figure 5 – Mean bacterial load in spontaneous, non-spontaneous PTB and control placentas (with 95% CIs)

Ongoing and Future Work

16s rRNA gene sequencing for species specific identification

- Library preparation for next generation sequencing of regions V5-V7 of bacterial 16s rRNA gene underway.
- Fetal placental and maternal villous samples will all be run on MiSeq sequencer by year end.
- Plan to sequence bacterial 16s DNA from cord blood and maternal blood, which will provide information relating to possible routes of bacterial colonization.
- Compare bacterial profiles of cohort placental tissues to those from Human Microbiome Project (HMP) – particularly vaginal and oral – to identify similarities and differences.

High throughput cytokine assay to identify maternal immune response

- Collaboration established with group at Virginia Commonwealth University who have recently received funding from HMP to conduct a pregnancy microbiome project (MOMS-PI).
- Plan to travel to lab in spring/summer next year to conduct high-throughput cytokine panel on maternal serum (and possibly cord blood) from preterms and controls to examine relationship between bacterial profiles observed through sequencing and maternal/fetal immune response.

References

2014. The Integrative Human Microbiome Project: Dynamic Analysis of Microbiome-Host Omics Profiles during Periods of Human Health and Disease. *Cell Host Microbe,* 16**,** 276-289.

Abu-Amero et al. 2014. The Baby Bio Bank-A Legacy for Researchers Worldwide into Common Complications of Pregnancy. *Journal of General Practice,* 2**.**