UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Genome-wide gene expression analysis in black South African women who develop gestational diabetes mellitus

Angela Wendy Hobbs

A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy

December 2017

Declaration

I, Angela Wendy Hobbs, hereby declare this thesis is my own, unaided work. It is being submitted for the degree of Doctor of Philosophy of Health Sciences at the University of the Witwatersrand, Johannesburg. It has not been previously submitted for any degree or examination at this or any other University.

26/08/2017

Angela Wendy Hobbs *Date*

Dedication

In loving memory of my father,

Wayne Dennis Hobbs

1955 – 2001

Acknowledgements

- To my supervisor and mentor, Professor Michèle Ramsay, thank you very much for your patience, constant encouragement and support over the years and for providing me with many invaluable opportunities and advice. I am forever grateful for your mentorship.
- To my co-supervisor, Professor Shane Norris, and to Professor Nigel Crowther, thank you for your help and advice.
- To Dr Xavier Estivill and the lab members at the Centre for Genomic Regulation (CRG) in Barcelona, thank you for your help with my project during my stay, I appreciate the time and effort you took to assist me.
- A big thank you to the principal investigators and students of the Non-Communicable Disease Research and Leadership (NCD) programme for the years of support and financial assistance.
- To my family for your constant love, support and encouragement along the way. Love you all very much.
- To my husband, Brendan, for allowing me to fulfil my dreams and encouraging me every step of the way and too my son, Zak, born on the 26th December 2015, I love you.
- To God, for the hand you play in every aspect of my life and giving me the strength to never give up.

Abstract

Gestational diabetes mellitus (GDM) is characterized by high blood glucose levels that first develop during pregnancy. GDM has been linked with many adverse short and long term health outcomes for the developing foetus as well as for the mother. The Developmental Origins of Health and Disease (DOHaD) concept suggests that in the presence of adverse stimuli, the foetus will adapt, through epigenetic mechanisms, to ensure its immediate survival. For this reason, epigenetic modifications are emerging as mediators linking early environmental exposures during pregnancy with programmed changes in gene expression that alter offspring growth and development. The objective of this research study was to explore the role of altered gene expression and methylation in the development of GDM and determine whether these alterations are inherited by the exposed foetus.

Transcriptome sequencing was performed on mRNA extracted from blood samples collected from six women with GDM and from six controls; as well as from exposed (N=6) and unexposed placenta (N=6). Genes that displayed significant ($p<0.005$) differential expression (log2 fold change >2 and <-2) between cases and controls were identified from the blood (N=60) and placenta (N=56) datasets. Gene ontology and enrichment was performed using DAVID and PANTHER with the aim to narrow down the candidate gene lists.

The ten most likely candidate genes for differential gene expression from the blood dataset were *G6PD, DCXR, TKT, ALDOA, PGLS, KCNQ1, C14orf80, KCNQ1, SLC25A22* and *GSK3A.* Gene enrichment revealed that five of these significantly under-expressed genes *(G6PD, DCXR, TKT, ALDOA* and *PGLS)* encode enzymes in the pentose phosphate pathway (PPP). In the placental dataset the top ten candidate genes were *CXCR1, CXCR2, G6PD, TKT, IGFBP-1, IGFBP-2, IGFBP-6, GGT3P, MMP12* and *GLT1D1*. The direction and fold change of differential expression of all twenty genes were validated using TaqMan qPCR probes. Of these twenty genes, the five most promising biological candidates *(G6PD, TKT, IGFBP-1, IGFBP-2* and *IGFBP-6)* were identified and the level of promoter region methylation was assessed using EpiTech Methyl II PCR Assays. The level of methylation in the promoter region of *G6PD* in both blood and placenta

tissue was found to be significantly higher ($p=1.90 \times 10^{-5}$ and $p=1.2 \times 10^{-11}$ respectively) in the case groups, correlating with decreased mRNA expression levels. There was a significant negative correlation between *G6PD* mRNA expression in the blood and placenta with the level of maternal glucose at fasting (p=0.006 and p=0.001, respectively), 1-hr (p=0.016 and p=0.007, respectively) and 2-hr post OG (p=0.045 in placenta). We observed a significant positive correlation between *G6PD* promoter region methylation in both blood and placental tissues with maternal glucose levels at fasting (p=0.023 and p=0.001, respectively) and at 1-hr post OG (p=0.001 and p=0.004, respectively). *IGFBP-1* was found to be significantly under-expressed in exposed placental tissue and hypermethylated ($p=1.1 \times 10^{-6}$) at the promoter region when compared to unexposed samples. There was a significant negative correlation between the expression of *IGFBP-1* mRNA in the blood and placenta with foetal birth weight (p=0.005 and p=0.017, respectively).

Our results suggest that high glucose levels, an important characteristic of GDM, result in the disturbance of the pentose phosphate pathway, a pathway linked closely to glycolysis, and the IGF-axis, which is important in foetal growth and development. In GDM there is suppression of *G6PD* mRNA expression in both the blood and placental tissue which influences the pentose phosphate pathway. We hypothesize that this is mediated through an epigenetic mechanism since it is correlated with increased methylation of the *G6PD* promoter region. Down regulation of *G6PD* would suppress the PPP and reduce the levels of NADPH production, which may in turn lead to an increase in oxidative stress and an adverse outcome in the mother and foetus. With regard to the IGF-axis, our results demonstrated that *IGFBP-1* and *IGFBP-2* mRNA expression in the placenta may be inhibited due to the presence of high glucose and insulin levels and this decrease in mRNA expression is likely implicated in the abnormal foetal growth which is often associated with GDM.

This study has provided novel insights into gene expression and DNA methylation changes in the blood of women with GDM and the placenta of their female offspring that involve genes in the PPP and the IGF-axis.

Abbreviations

ACTB	Beta Actin
ADA	American Diabetes Association
ALDOA	Fructose-biphosphate aldolase A
ATP	Adenosine Triphosphate
bp	Base pair
BMI	Body Mass Index
C	Cytosine
cDNA	Complementary DNA
CDS	Coding DNA Sequence
cm	Centimetre
CO_2	Carbon dioxide
CRG	Centre for Genomic Regulation
C_T	Threshold Cycle
CVD	Cardiovascular disease
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DCT	Diluted Cluster Template
DCXR	Xylulose Reductase
DEC	Dependant Enzyme Control
Δ	Delta
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA Methyltransferases
DOHaD	Developmental Origins of Health and Disease
DPHRU	Developmental Pathways and Health Research Unit
dsDNA	Double stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EWAS	Epigenome-wide association studies
FDR	False Discovery Rate
FGR	Foetal Growth restriction
FPKM	Fragments per kilo base of transcript per million mapped
g	Gram
G	Guanine
GC	Guanine-cytosine
GDM	Gestational Diabetes Mellitus
G6PD	Glucose-6-phosphate dehydrogenase
GSEA	Gene set enrichment analysis
HAPO	Hyperglycaemia and Adverse Pregnancy Outcome
hbA1c	Glycated haemoglobin
HPRT1	Hypoxanthine phosphoribosyltransferase 1
hr	Hour
IADPSG	International Association of Diabetes and Pregnancy Study Group
IAP	Intracisternal A particle
IGFBP	Insulin-like growth factor binding protein
IGF	Insulin growth factor
Insr	Insulin receptor
IUGR	Intrauterine growth restriction
kb	Kilobase
kDA	KiloDalton
kg	Kilogram
LCL	Lymphoblastoid cell line
LGA	Large gestational age
Line-1	Long interspersed nuclear elements
MDRE	Methylation dependant restriction enzyme

MEST	Mesoderm specific transcript
meQTL	Methylation quantitative trait loci
mg	Milligram
mg/dL	Milligrams per decilitre
ml	Millilitre
mmol	Millimol
mRNA	Messenger RNA
MSRE	Methylation Sensitive Restriction Enzyme
N	Number
NaCl	Sodium Chloride
NADH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Centre for Biotechnology Information
NCD	Non-communicable Disease
NGS	Next generation sequencing
NGT	Normal Glucose Tolerance
ng	Nanogram
nm	Nanomolar
NRT	No reverse transcriptase control
OGTT	Oral glucose tolerance test
OG	Oral Glucose
OMIM	Online Mendelian Inheritance in Man
PANTHER	Protein Analysis through Evolutionary Relationships
PBS	Phosphate buffered saline
PCR	Polymerase Chain reaction
PDP	Pooled DCT plate
Pdx1	Pancreatic/duodenum homeobox protein 1
PEG3	Paternally expressed gene 3
PGLS	6-Phosphogluconolactonase
Pomc	Proopiomelanocortin
Ppara	Peroxisome proliferator activated receptor alpha
PPP	Pentose phosphate pathway
pSoBid	Psychological, social and biological determinants of ill health
QC	Quality Control
qPCR	Quantitative polymerase chain reaction
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
ROS	Reactive oxygen species
RPKM	Reads per kilobase of transcript per million mapped
RPLPO	Large Ribosomal Protein
rRNA	ribosomal RNA
RQ	Relative Quantification
SDS	Sodium dodecyl sulphate
SEC	Sensitive Enzyme Control
SES	Socio economic status
SGA	Small gestational age
SNP	Single Nucleotide Polymorphism
SSA	Sub-Saharan Africa
TE	Tris-EDTA
T2D	Type 2 diabetes
T1D	Type 1 diabetes
µl	Microliter
TKT	Transketolase
UCSC	University of California Santa Cruz
USA	United States of America
vs	Versus

WHO	World Health Organization
Wits	University of the Witwatersrand

Table of Contents

List of Figures

List of Tables

CHAPTER ONE

1. INTRODUCTION

1.1 Diabetes Mellitus in Sub-Saharan Africa (SSA)

The incidence of diabetes, especially type 2 diabetes (T2D), is rapidly growing worldwide. In 1980, an estimated 108 million people suffered with this chronic disease, which by year 2014, had increased to an estimated 422 million, representing 8.5 % of the worlds adult population. Of this number, 80 % of these individuals live in developing countries (Roglic et al., 2005, Azevedo and Alla, 2008). Diabetes was once considered a rare disease in SSA however, like the rest of the world, SSA is experiencing an increasing prevalence of this disease. It is predicted that SSA will have the highest growth in the number of people affected by diabetes over the next 20 years. The prevalence of diabetes in SSA is projected to rise from 12.1 million in 2010 to an estimated 23.9 million in 2030, representing a 98 % increase in the number of individuals affected by the disease (Shaw et al., 2010, Tekola-Ayele et al., 2013). This proportion is more than double the predicted global increase of 37 % (Mbanya et al., 2011). Assessing the prevalence and incidence of diabetes in SSA is challenging because of the lack of data from many countries (Mbanya et al., 2011). The majority of the African diabetes is of type 2 (70 %–90 %), with only 25 % showing the complications of type 1 (Mufunda et al., 2006, Osei et al., 2003). Although there is a strong genetic predisposition to developing T2D, the alarming increase in its prevalence in SSA is mainly attributed to changes in lifestyle and diet with increased food intake and reduced energy expenditure (Alfaradhi and Ozanne, 2011, Mbanya et al., 2011). The process of urbanisation is occurring at a rapid rate in SSA and has a large influence on the prevalence of diabetes. The relocation to urban areas results in lifestyle changes as well as changes in dietary habits (Beaglehole and Yach, 2003). The prevalence of diabetes in urban residents is almost 4 times higher than in rural dwellers (Jamison et al., 2006) and this number is expected to increase substantially due to a high rate of urbanisation.

Diabetes is a chronic disease that requires life-long treatment and significantly increases the risk of serious, long-term complications (Hall et al., 2011, Mossie et al., 2017). The rapid increase in the prevalence of diabetes in SSA poses a major and costly public health and socioeconomic burden. Offering the long-term monitoring and treatment needed is not easy

for the healthcare systems of SSA which are more concerned with managing severe infections (Azevedo and Alla, 2008). There are many factors which limit the appropriate diabetes care in SSA, namely inadequate healthcare systems, the shortage of adequately trained doctors and nurses and the unaffordability of medication (Mbanya et al., 2011). On an individual level, those affected with diabetes face many practical and financial problems. Not only is it difficult to reach treatment centres, but the necessary medication is expensive and often not affordable to many families affected with diabetes. With the double burden of both infectious and non-communicable disease in the SSA region, diabetes must compete for political attention and financial investment (WHO, 2013). It is possible to prevent this burden if effective interventions are implemented. Although the incidence of diabetes can be reduced by primary prevention and treatment, establishing effective diabetes programs in SSA, a region that requires a shift in the current public health priority, requires a larger amount of evidence to highlight the magnitude of the problem as well as the areas for intervention (Hall et al., 2011). Based on aetiology, diabetes mellitus has been classified into type 1 diabetes (T1D), type 2 diabetes (T2D) and gestational diabetes mellitus (GDM). T1D is the most studied type of diabetes and is characterised by the autoimmune destruction of pancreatic B-cells leading to insulin deficiency (Evangelista et al., 2014). Macrophages, dendritic cells and lymphocytes are involved in this pathogenic process through a complex interplay of mechanisms implicated in the loss of immune tolerance to autoantigens. T2D, a genetically heterogeneous disease with several rare monogenic forms and a number of common forms resulting from a complex interaction of genetic and environmental factors, is characterised by insulin resistance and pancreatic ß-cell dysfunction (Doria et al., 2008). T2D arises from an impairment in the ability of muscle, fat and liver to respond to insulin, combined with an inability of the ß-cells to respond normally to glucose by increasing insulin secretion (Kahn, 1994). GDM is a complication of pregnancy that is characterised by impaired glucose tolerance with the onset or first recognition during pregnancy (Al-Badri et al., 2015).

1.2 Gestational diabetes mellitus (GDM)

1.2.1 Definition

Diabetes can affect pregnancy in two ways; pre-gestational and gestational diabetes. Women diagnosed with type 1 diabetes (T1D) or T2D prior to conception are referred to as having pre-

gestational diabetes but when pregnant women exhibit high fasting blood glucose levels first observed during pregnancy, it is referred to as gestational diabetes mellitus (Ben-Haroush et al., 2004, Jawad and Ejaz, 2016). GDM is defined as “any degree of glucose intolerance with onset or first recognition during pregnancy” (Durnwald, 2015, Karagiannis et al., 2010). GDM is characterized by β-cell function that is insufficient to meet the increased demand for insulin during pregnancy. The severity and prevalence of GDM, a heterogeneous disorder, is influenced by a number of different factors including genetic background, obesity and age (Butte, 2000). Although GDM resolves after delivery, uncontrolled GDM is associated with an increased risk of maternal and neonatal morbidity and predisposes the mother as well as the foetus to a higher risk of developing T2D, obesity and metabolic syndrome later in life. There are some striking parallels between GDM and T2D; GDM shares several risk factors with T2D and the pathophysiological changes of GDM and T2D are similar, both resulting by insulin resistance accompanied by an insulin-secretory defect (Buchanan, 2001). There is an observed association between GDM and the risk of developing T2D later in life (Eades et al., 2015, Kaaja and Ronnemaa, 2008, Kim et al., 2012, Noctor and Dunne, 2015). This suggests that GDM may serve as a window for determining predisposition to T2D (Robitaille and Grant, 2008).

1.2.2 Pathophysiology

Insulin resistance is one of the pathophysiological mechanisms which underlie the development of GDM (Catalano et al., 2003). The specific mechanisms underlying the development of the disease are unknown. It is believed that the secretion of pregnancy hormones (specifically, estrogen, human placental lactogen, cortisol and progesterone) (Figure 1.1) interferes with the action of insulin as it binds to the insulin receptor. This interference will result in a state of insulin resistance and an accumulation of glucose in the blood. Insulin resistance is a natural phenomenon occurring in all pregnancies to ensure that a sufficient amount of glucose reaches the developing foetus. To overcome this increase of glucose, more insulin secretion is required. However, in a certain percentage of women, the up-regulation of insulin secretion does not occur and glucose accumulation will continue to levels seen in non-pregnant T2D individuals (Gabbe, 1986, Poulakos et al., 2015). Maternal hyperglycaemia, foetal hyperinsulemia and foetal over nutrition is the result of this imbalance

between increased insulin resistance and maternal insulin production observed in GDM pregnancies. Although a number of explanations regarding the inability to regulate insulin needs during pregnancy have been proposed, the reason remains unknown. The clinical presentation of GDM is well defined however, the mechanism underlying the development of this disease is still not well understood (Catalano et al., 2003; Hajj et al., 2014).

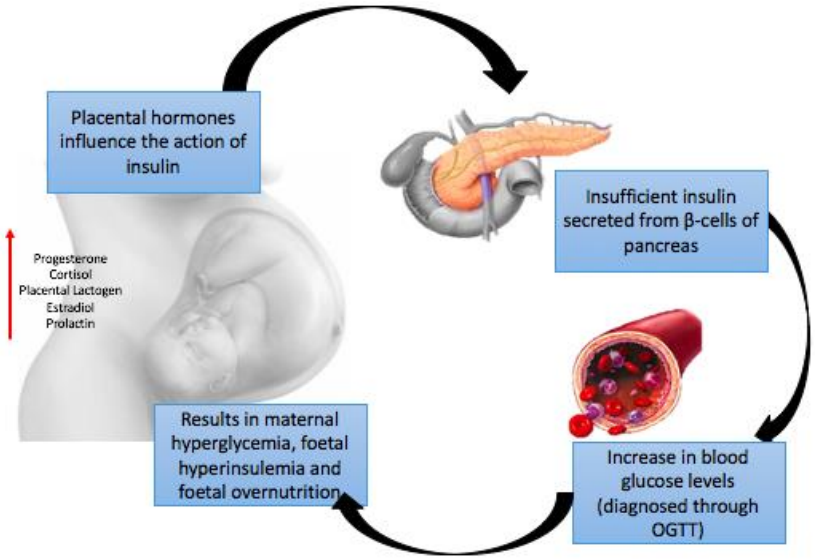


Figure 1.1. The proposed pathophysiology of gestational diabetes. Placental hormones partially block the action of insulin which results in insulin resistance in pregnant women. Normally, there is an up-regulation of insulin secretion from ß-cells to restore this insulin resistance, however in a certain percentage of women, this up-regulation does not occur and glucose will accumulate, resulting in GDM.

1.2.3 Prevalence

Gestational diabetes mellitus is characterized by glucose intolerance that first becomes apparent during pregnancy. GDM is becoming a major public health concern as its prevalence has doubled over the past 20 years, affecting approximately 16.9 % of pregnancies worldwide (Arora et al., 2013, Gilmartin et al., 2008, Robitaille and Grant, 2008). Macaulay et al. (2014) reviewed data from 14 studies conducted in six African countries and reported the prevalence of GDM to range from 0 % in Tanzania to 13.9 % in women from Nigeria (Macaulay et al., 2014). A more recent review reported data from 22 studies in six African countries and observed that, regardless of the diagnostic criteria and study setting, the prevalence of GDM in SSA is in a range comparable to the 2-6 % reported for European countries (Mwanri et al., 2015). These data show that the prevalence of GDM varies depending on the diagnostic

criteria used and also between racial and ethnic groups (Buchanan, 2001, Buchanan and Xiang, 2005, Noctor and Dunne, 2015). Studies that compared different diagnostic criteria found that there were significant differences in the reported prevalence of GDM in SSA (Mwanri et al., 2015), however, in line with an increasing prevalence of T2D, the incidence of GDM in SSA is also on the rise (Petry, 2010).

1.2.4. Risk Factors

There are a number of different factors that will increase a woman's risk of developing GDM (Bottalico, 2007, Zhang et al., 2016). The risk factors for GDM share similarities with those for T2D with the common factors including a previous history of GDM or impaired glucose tolerance (Zhang and Ning, 2011), obesity, ethnicity (higher risk for those of African, Asian, Hispanic, and Native American descent), advanced maternal age (older than 35 years), a family history of T2D (Zhang and Ning, 2011), a history of spontaneous abortions and unexplained stillbirths as well as polycystic ovarian syndrome and pregnancy related hypertension (Gilmartin et al., 2008, Petry, 2010).

The presence of GDM has been linked with short and long term risk factors for both the mother and her offspring (Enquobahrie et al., 2009). Short-term risks associated with GDM in the mother include a higher incidence of pre-eclampsia (Oats and Beischer, 1986), unexplained stillbirths and spontaneous abortions, urinary tract infections, increased risk of congenital abnormalities as well as an increased risk of postpartum bleeding, pregnancy-induced hypertension and abnormal weight gain. GDM has long been known to increase the risk of macrosomia (birth weight > 4 kg) in the offspring (Petry, 2010). GDM-associated macrosomia is linked to increased rates of a variety of complications, including shoulder dystocia, brachial plexus injuries and clavical fractures which occur during natural birth, leading to a higher requirement for Caesarean section (Catalano et al., 2003, Petry, 2010, Watanabe, 2011). Diabetic pregnancy induces marked abnormalities in glucose homeostasis and insulin secretion in the foetus, resulting in foetal hyperglycaemia and abnormal foetal growth (Lehnen et al., 2013, Pinney and Simmons, 2012, Vambergue and Fajardy, 2011). Short-term risks for the baby include neonatal hypoglycaemia (Watanabe, 2011); neonatal cardiac dysfunction; respiratory stress disorder; hyperbilirubinemia and hypocalcaemia

(Petry, 2010, Robitaille and Grant, 2008). Other early effects of GDM on the foetus include jaundice, polycythemia, hypocalcemia, hypomagnesemia, and prematurity (Yang et al., 2002).

Women who develop GDM have an increased risk (ranging from 17 % - 63 %) of developing T2D, within 5 - 16 years after pregnancy, depending on population group and other risk factors (Robitaille and Grant, 2008). As well as long-term risks for herself, her offspring will have increased risk of developing childhood obesity, T2D and hyperlipidaemia in adolescence as well as developing adult onset disease such as cardiovascular disorders (Law et al., 2015, Pettitt et al., 2008). When the offspring is female, exposure to maternal diabetes *in utero* increases their risk of developing GDM in their own pregnancies (Claesson et al., 2007). This metabolic programming by *in utero* exposure to hyperglycaemia is a transgenerational effect that may contribute to the large increase in the prevalence of T2D worldwide (Petry, 2010).

1.2.5 Screening and Diagnosis

Gestational diabetes mellitus, a condition that is generally asymptomatic, is only diagnosed if screening is done during pregnancy (Buchanan et al., 2012). Despite the increasing evidence showing that untreated GDM is associated with short and long term risks to the mother and her offspring, controversy remains regarding screening tests, diagnostic tests and the level of hyperglycaemia that is diagnostic of GDM (Gilmartin et al., 2008). There are two primary methods used to screen for GDM, a universal approach method and a risk factor based method (Figure 1.2)(Gilmartin et al., 2008, Rani and Begum, 2016). The universal screening approach maximizes sensitivity and will identify women with GDM who have no risk factors whereas the risk factor based approach is more cost effective due to its selective screening of women who have the most common GDM risk factors (Berger and Sermer, 2009).

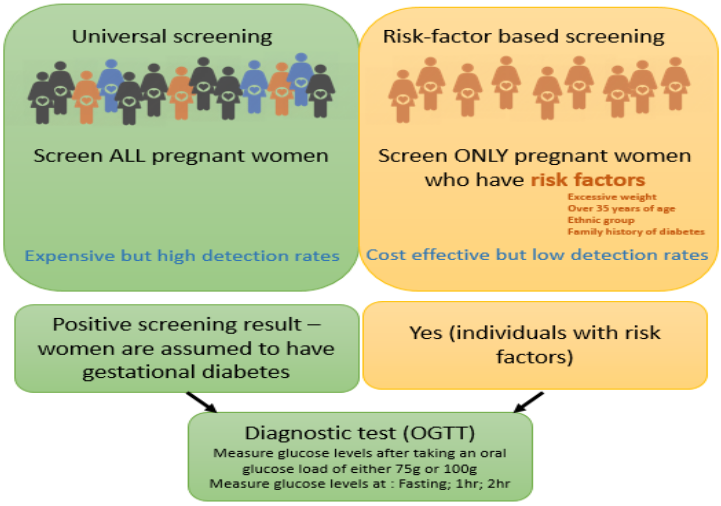


Figure 1.2: Universal and risk factor based strategies for screening pregnant women for GDM. The most common risk factors include family history of the disease, excess weight, ethnic groups and age greater than 25 years.

Diagnostic tests differ from the above-mentioned screening tests. GDM is usually diagnosed using some kind of Oral Glucose Tolerance Test (OGTT). During the years there have been different diagnostic criteria for the diagnosis of GDM but all provide diagnostic cut-off levels for diabetes using blood glucose concentrations (Schneider et al., 2003). There are a number of different diagnostic criteria used to classify diabetes during pregnancy. O'Sullivan and Mahan proposed the first criteria for the diagnosis of GDM in 1964 (O'Sullivan and Mahan, 1964). These criteria were later modified by Carpenter and Couston (Carpenter and Coustan, 1982). Both these early criteria focused mainly on the mother's postpartum risk of developing diabetes because the adverse risk to the foetus was not yet known. The World Health Organization (WHO) criteria from 1985 are still widely used throughout the world for the diagnosis of GDM (WHO, 2013), however the more recently published guidelines have revised the diagnostic value thresholds that are indicative of GDM. The American Diabetes Association (ADA) recommendations are based on the Carpenter-Couston 100 g OGTT criteria (Karagiannis et al., 2010) and are used to screen all pregnant women in the USA. Based on the Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) study published in 2008 (Metzger et al., 2008), new recommendations were made by the ADA (Couston et al., 2010). This multicentre, multinational observational study of 25,000 pregnant women aimed to identify a relationship between maternal hyperglycaemia (less severe than overt diabetes) and adverse pregnancy outcomes. They found that the risk for macrosomic babies, neonatal

hypoglycaemia, increased C-peptide levels and caesarean delivery increased with an increase in the mother's glucose levels, even if they were below the value for GDM (Lowe et al., 2012, Metzger et al., 2008). Since this study, the International Association of Diabetes and Pregnancy Study Group (IADPSG) recommended that all pregnant women without known diabetes undergo a 75 g OGTT at 24 - 28 weeks gestation and that their fasting glucose, 1-hour and 2-hour glucose levels be considered important in the diagnosis of GDM (Bernasko, 2016). Using HbA1c in screening for GDM (instead of an OGTT) has been studied but was found to be controversial and resulted in misclassification and misdiagnosis (Salmeen, 2016). However, recent studies have confirmed that the HbA1c test can be used for the diagnosis of GDM but in conjunction with the OGTT (Aldasouqi et al., 2008, Soumya et al., 2015). Therefore, the OGTT remains the gold standard for GDM diagnosis. However, there are still no universal recommendations for the ideal approach for the screening and diagnosis of GDM today. With the number of women who develop GDM during pregnancy increasing at an alarming rate (Liao et al., 2012) and the increasing evidence that GDM is associated with adverse pregnancy outcomes for mothers and their offspring (Kalter-Leibovici et al., 2012), screening of all pregnant women should be done routinely.

1.2.6 Treatment and prevention

Although there are no evidence-based studies which indicate that the prevention or treatment of GDM minimize maternal or foetal complications (Jovanovic, 2001, Jovanovic and Pettitt, 2001), there have been a number of studies done which aim to demonstrate the effect treatment has on GDM (Bancroft et al., 2000, Elnour et al., 2008, Nachum et al., 1999). Nutritional intervention for women with GDM has been recognised as the cornerstone of management therapy (Funnell et al., 2007), although other treatment types include interventions to control blood glucose, diet control, glucose monitoring, insulin use and pharmaceutical intervention (Horvath et al., 2010, Kalter-Leibovici et al., 2012). These studies demonstrated that these treatment types did significantly reduce the risk of adverse perinatal and neonatal outcomes (Kalra et al., 2016, Tobias et al., 2011, Tobias, 2011) and highlight that the prevention of GDM is crucial in avoiding the adverse outcomes often associated with GDM. Physical activity is known to improve glucose homeostasis through its direct and indirect effects on insulin sensitivity (Colberg et al., 2013). Physical activity has independent effects on

glucose disposal by increasing insulin mediated and non-insulin mediated glucose disposal (Tobias et al., 2011, Tobias, 2011). The long-term effects of physical activity include a significantly improved insulin sensitivity through increased fat free mass. However, more data are required to significantly show the impact physical activity in women with GDM has in the prevention of, or delay in, the development of T2D (Colberg et al., 2013, Tobias et al., 2011). A review written in 2015 (Sanabria-Martinez et al., 2015) found that moderate physical exercise done during pregnancy was effective for the prevention of GDM, however, a study done by Yin et al. (2014) did not find a significant effect (Yin et al., 2014) suggesting that the effect physical activity has on GDM and the consequent development of T2D remains controversial.

1.3. Developmental programming of adult disease

Diabetes is a complex trait that results from interactions between genes, dietary intake, physical inactivity and other environmental factors. Although a number of genes have been known to play a role in the development of diabetes, the genetic component alone cannot account for the dramatic increase in the prevalence of this disease and environmental factors must be important triggers (Vickers, 2011). Therefore, the development of diabetes, a multifactorial disease, is influenced by genetic and environmental factors. Environmental factors which have the ability to modify genetic risk are important in the development of diabetes (Murea et al., 2012). Epigenetics may be the molecular link between environmental factors and diabetes. These environmental factors may cause epigenetic changes that can persist into adulthood. Relevant epidemiological studies have observed an association between the *in utero* environment and the subsequent development of adult disease. The terms “developmental programming” and the “developmental origins of adult health and disease (DOHaD)” are used to describe these associations (Charles et al., 2016).

1.3.1 The theories of developmental programming

It is now well recognized that the *in utero* environment influences key developmental processes and has long-lasting effects on health and disease (Hajj et al., 2014). In 1986, Freinkel proposed the ‘fuel-mediated teratogenesis’ concept which postulates that foetal hyperinsulemia in pregnancies affected by GDM, is a result of the increase in glucose (Freinkel

et al., 1986). This has immediate as well as long-lasting consequences for the offspring (Hajj et al., 2014). In the early 1990s, Barker and Hales (1992) proposed the concept of foetal metabolic programming which stated that "foetal and perinatal events, such as maternal over and under-nutrition, were crucial in determining the risk of developing chronic metabolic diseases in adulthood" (Hales and Barker, 1992).

The thrifty hypothesis describes how intrauterine malnutrition influences foetal development and increases the offspring's risk of developing adult disease (Chen and Zhang, 2011). This hypothesis suggests that the exposure to a nutritionally suboptimal *in utero* environment will result in a permanent alteration of glucose metabolism, thereby increasing the risk of developing T2D in adulthood. In 1993, Barker et al. (1993) hypothesized that neonates born with low birth weight, a marker for an adverse *in utero* environment, had increased susceptibilities for developing metabolic diseases in adulthood (Barker et al., 1993, El Hajj et al., 2014b, Hanson et al., 2004). This hypothesis is commonly referred to as the DOHaD hypothesis and is referred to in more detail in Section 1.3.2. Keeping the perspective of DOHaD in mind, Hanson and Gluckman (2004) proposed the Predictive Adaptive Response hypothesis, which states that "foetal reprogramming, induced by *in utero* exposures, is a short term adaption to the predicted environment in order to enhance the survival of the individual". It is hypothesized that these adaptions may occur through epigenetic changes (Gluckman and Hanson, 2004, Ueda, 2013). In this way, the foetus is predicting what its early life nutritional environment will be like based on its *in utero* exposure. Although this adaption may improve the chances of survival in the short term, they may be deleterious to long term health given a more nutrient rich environment (Hales and Barker, 1992, Ueda, 2013). The problem of "mismatch" occurs when individuals adapt to one environment during development and are then exposed to another after birth. The association observed between increased susceptibility to developing adult onset diseases and nutritionally adverse *in utero* environments is of particular interest in low and middle income African countries (Hobbs and Ramsay, 2015). In these countries, people are born into rural environments that are nutrient scarce and move to urban areas where they are exposed to unhealthy diets, smoking, drinking, and become less active (all NCD risk factors). The rapidly increasing burden of disease in these developing economies may be underpinned by *in utero* exposure to nutrient poor

environment (Puoane, 2008, Ueda, 2013). This phenomenon is thought to be involved in the current T2D epidemic (Brenseke et al., 2013). The current surge in metabolic and cardiovascular disease in SSA may be fueled by a combination of under-nutrition in early life and over-nutrition in later life.

1.3.2 The DOHaD hypothesis

It is well known that the phenotype of an individual can be determined by both the *in utero* and early postnatal environmental conditions such as the nutritional state of the mother (Alfaradhii et al., 2011). Barker and colleagues observed that the starvation of pregnant women during the Dutch 'hunger winter' of the second world war correlated with an increased risk of cardiovascular and metabolic diseases in their offspring later in life (Barker et al., 1993, Heindel and Vandenberg, 2015). Hence, a link between early life environmental factors and later life diseases was observed. These observations opened a new and exciting area of research. DOHaD is a multi-disciplinary field that studies how "environmental factors acting during the phase of developmental plasticity interact with genotypic variation to change the capacity of the organism to cope with its environment in later life" (Gluckman and Hanson, 2004, Uauy et al., 2011). Substantial experimental evidence from different mammalian species has supported the hypothesis that an adverse intrauterine environment may alter the embryo's development through predictive adaptive responses. It is believed that these responses are adaptations made by the embryo in response to a particular maternal environment to ensure its immediate survival *in utero* and to prepare itself for postnatal life. It is likely that a disease phenotype will arise if there is a mismatch between the *in utero* diet and postnatal diet (Trussler, 2010). Evidence from epidemiological studies as well as animal models suggest that the intrauterine environment does play a role in the development of metabolic disorders. A suboptimal maternal nutrition, whether under- or over-nutrition, has negative effects on the offspring (Williams et al., 2014).

1.3.2.1 Evidence from human epidemiological studies for maternal nutrition influencing early life exposure to disease risk

Two separate studies examined the relationship between glucose concentrations at 28 weeks gestation and the development of early childhood obesity. Deierlein et al. (2011) observed a

two-fold greater risk of childhood overweight/obesity at 3 years of age when comparing maternal glucose concentrations ≥130 mg/dL to maternal glucose concentrations less than 100 mg/dL during GDM testing. These results indicate that foetal exposure to high glucose concentrations may contribute to the development of overweight/obesity in the offspring (Deierlein et al., 2012, Kim et al., 2012). Pettit et al. (2010) evaluated the relationship of glycemic levels during pregnancy with anthropometric data of the offspring from diabetic and nondiabetic mothers of the HAPO study. They found a significant association with maternal 1-hour blood glucose levels and offspring BMI at two years of age. Crume et al. (2011) examined the association between exposure to maternal diabetes *in utero* and BMI growth trajectories from birth through age 13 years. After adjusting for sex and ethnicity, they observed that the overall BMI growth trajectory was not significantly different between exposed and unexposed offspring from birth to 26 months. However, from the age of 27 months to 13 years, the BMI growth trajectory for exposed offspring was significantly greater than that of unexposed offspring (Crume et al., 2011).

Following the initial work of Barker et al. (1993) that demonstrated a relationship between low birth weight and an increased risk of adult disease, the importance of maternal nutrition was addressed in studies of famine exposure (Vickers, 2011). The Dutch Hunger Winter study demonstrates the long-lasting effects that nutritionally adverse *in utero* and/or neonatal environments have on health and disease (Lehnen et al., 2013). This Dutch Hunger Winter cohort included men and women who were exposed *in utero* to the Dutch famine of 1944-1945 (Heijmans et al., 2008, Lehnen et al., 2013). Individuals exposed to a nutritionally adverse (under-nutrition) *in utero* environment exhibited an increased risk for metabolic diseases (El Hajj et al., 2014a). They also observed that women who were pregnant during this period and therefore exposed to environmental stress and nutrient restriction throughout their pregnancy, gave birth to infants who had a decreased birth weight and increased insulin resistance (Lumey et al., 2007) when compared to their unexposed siblings.

In rural Gambia, there is seasonal variation in the availability of micronutrients with an alternation between the dry season (when food is plentiful) and the wet season (when there is less food available and therefore poorer nutrition) (Waterland et al., 2010). During the

nutritionally poor rainy season, a high incidence of deficiencies in several essential micronutrients has been observed in pregnant women (Waterland et al., 2010). This seasonal deficiency is associated with an increased incidence of low birth weight, as well as childhood morbidity and mortality (Khulan et al., 2012, Waterland et al., 2010). Another famine study which demonstrates the long-lasting effects nutritionally adverse *in utero* and/or neonatal environments have on health and disease is the Nigerian (Biafran) famine of 1967 – 1970 (Hult et al., 2010). The Biafran famine cohort consists of 1,339 Igbo individuals who were born before, during and after the Biafran famine (between 1965 and 1973). Hult et al. (2010) observed an increased risk of adult hypertension, glucose intolerance and obesity in individual's exposed to the famine *in utero* or in infancy (Hult et al., 2010). This study highlights that the prevention of under-nutrition during pregnancy and in infancy must become a high priority in health, education, and economic agendas (Hult et al., 2010, Ueda, 2013).

1.3.2.2. Evidence from animal studies

Animal models have been used extensively to study the physiological principles of the DOHaD hypothesis and are essential in studying the mechanistic links between prenatal and postnatal influences and the risk for developing the metabolic syndrome in later life (Vickers, 2011). A large number of studies have used dietary restriction (such as a protein and/or calorie restriction) as a model for observing the effects of maternal under-nutrition. A study performed by Yura et al. (2005) showed that offspring of dams fed a relatively modest 70 % nutrient restricted diet developed obesity and adiposity when compared to controls. In 2004, Bellinger and colleagues observed that the offspring of dams fed a 50 % protein restricted diet showed an increased appetite for energy dense food in early life when compared to controls (Bellinger et al., 2004). Although initial studies focused mainly on the effects of maternal under nutrition such as diet low in protein or calories, recent studies have focused on the effects of maternal over-nutrition (maternal high fat diet). In one such study, Samuelsson et al. (2008). Observed that the offspring of dams fed a palpable obesogenic diet exhibited increased adiposity as well as cardiovascular and metabolic dysfunction when compared to controls (Samuelsson et al., 2008). A study done later by the same researchers found that feeding sucrose to a mouse during pregnancy leads to hypertension and insulin resistance in female offspring (Samuelsson et al., 2013). These studies suggest that an adverse *in utero* nutritional

environment programs the development of a metabolic syndrome-like phenotype (Rinaudo and Wang, 2012).

1.4. Epigenetic mechanisms in developmental programming of adult disease

Although there have been several hypotheses proposed to explain the associations between *in utero* environment and adult health (Rinaudo and Wang, 2012), it is still is not well understood how foetal developmental plasticity enables organisms to make adaptive responses to the foetal environment that can result in permanent adverse effects later in life (Koukoura et al., 2012). Recent studies have suggested that the environment influences epigenetic processes that regulate gene expression patterns, and might play key roles in the developmental programming of adult disease (Chen and Zhang, 2011). The ability of the organism to change its structure and cellular function in response to the environment is known as developmental plasticity which, in turn, is known to act through epigenetic changes in gene transcription, alterations in tissue differentiation and changes in homeostatic processes (Gluckman and Hanson, 2004, Liguori et al., 2010). These epigenetic changes are established early on in life and control the expression of certain genes during development. Disease arises from a mismatch between deprived *in utero* environment (for which epigenetic adaptions have been imposed in early development) and a comparatively rich postnatal environment (Rinaudo and Wang, 2012). The plasticity of the developmental process allows the organism to respond to the surrounding environment during early development when cells are differentiating and tissues developing. Although this plasticity allows the organism to adapt to changing environments, interference with these developmentally adaptive processes may have adverse effects on functions and increase risk for disease later in life (Barouki et al., 2012). It is during the developmental stage that epigenetic marks undergo critical modifications. Once a tissue or biological system is fully developed, it is less sensitive to environmental stimuli. Therefore the most sensitive period for epigenetic effects is different for each tissue and may extend into childhood or perhaps puberty or beyond (Barouki et al., 2012). Epigenetic changes that occur as a result of maternal diabetes may predispose the offspring to develop metabolic disease in adulthood. These epigenetic signatures of adverse

environmental exposure may be transmitted to successive generations. This cycle could contribute to the worldwide metabolic disease epidemic.
(El Hajj et al., 2014).

1.4.1 Introduction to epigenetics

The term epigenetics refers to the "stable and heritable patterns in gene expression that do not involve alterations in DNA sequence" (Cazaly et al., 2015). There are three main epigenetic mechanisms, namely, DNA methylation, histone modification and RNA interference (Moore et al., 2013). These epigenetic modifications are important for normal development and differentiation of distinct cell lineages in the adult organism (Handy et al., 2011). DNA methylation is a post-replication modification that is predominantly found in the cytosines of the dinucleotide sequence cytosine-phosphate-guanine (CpG) (Brenseke et al., 2013). In addition to DNA methylation, there are proteins which are associated with the organisation of DNA into nucleosomes. These proteins are known as histones and are subject to a large number of post-translational modifications (methylation, acetylation and/or phosphorylation) which control the structure and/or function of the chromatin (Brenseke et al., 2013, Sadakierska-Chudy and Filip, 2015).

Epigenetic modifications are flexible and can be modified by external influences (Handy et al., 2011). They also provide a mechanism that ensures the stable propagation of gene activity from one generation of cells to the next (Bollati and Baccarelli, 2010). Failure and/or improper reprogramming of the epigenetic machinery has been implicated in a broad range of diseases. Epigenetic modifications can be cell, tissue and sex specific as well as time dependant (Barouki et al., 2012). Each of the >200 cells types in the body have a specific combination of silenced and expressed genes, which are established during development and differentiation and stably inherited during cell divisions (Lehnen et al., 2013). Epigenetic modifications are influenced by several factors such as the environment, lifestyle, age and disease state and are not only sensitive to nutrients and physiological factors, but also to drugs, tobacco smoke, alcohol, industrial chemicals and other environmental exposures (Figure 1.3) (Barouki et al., 2012).

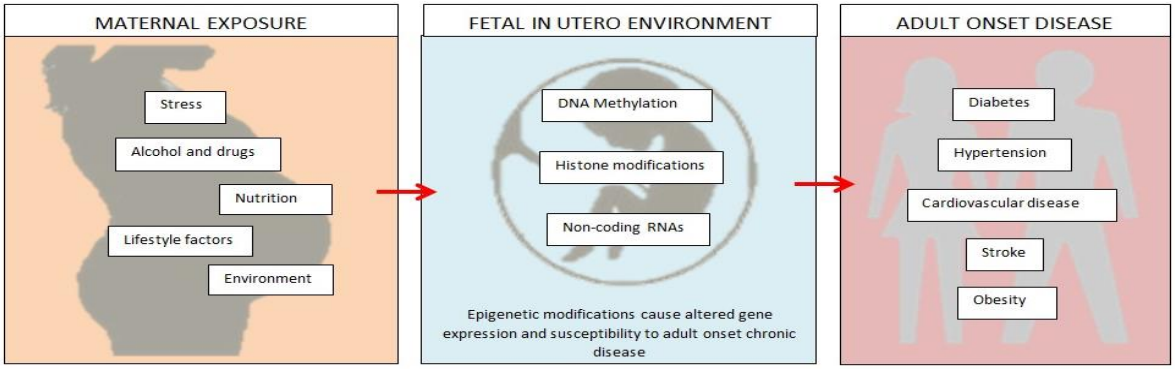


Figure 1.3: The link between an adverse *in utero* environment and epigenetic modifications. These epigenetic modifications take place in the developing foetus and may have lifelong consequences (taken from (Hobbs and Ramsay, 2015)).

1.4.2 DNA methylation

Since DNA methylation has the ability to modulate gene expression, which is a major determinant of many diseases, it has been suggested as a possible mechanism through which the exposure to an adverse *in utero* environment translates into the development of diseases (Alfaradhi and Ozanne, 2011, Waterland et al., 2010). DNA methylation appears to be the most dominant and best-studied epigenetic modification (Chen and Zhang, 2011) and is a heritable yet reversible epigenetic mark that can be stably propagated following DNA replication (Jin et al., 2011). DNA methylation involves the addition of a methyl group to the fifth carbon in CpG sites (Moore et al., 2013). The process of donating a methyl group to the 5-cytosine residue is catalysed by several enzymes known as DNA methytransferases (DNMTs). In humans, DNA methylation patterns are "established by the DNA methyltransferase DNMT3 family of de novo methyltransferases and maintained by DNMT1" (Law and Jacobsen, 2010).

In the non-coding regions of the genome, the majority of CpG dinucleotides are methylated in order to prevent retrotransposition activity. Unmethylated CpGs are usually clustered together in 'CpG islands', which are located within the promoter region of genes (Bird, 2002, Hill et al., 2011). In normal cells, the promoter CpG islands are typically unmethylated and are associated with active gene expression during differentiation and development. The methylation of these CpG islands during development or disease processes is associated with post-translational histone modifications. This modification results in a condensed inactive

chromatin structure and ultimately gene silencing (Lehnen et al., 2013). There are other regions of DNA methylation and intermediate CpG densities which exist across the genome, often in the body of genes. CpG island 'shores' are regions of comparatively low CpG density, located approximately 2 kb from CpG islands and exhibit tissue- and cancer-specific differential methylation (Fan and Zhang, 2009, Fernandez et al., 2012). Beyond CpG islands and shores, the remainder of genome displays a lower frequency of CpG sites (Stirzaker et al., 2014).

DNA methylation is crucial for normal mammalian development and is an important component for many cellular processes such as "embryonic development, genomic imprinting, X-chromosome inactivation and preservation of chromosome stability" (Sadikovic et al., 2008). DNA methylation typically represses transcription by inhibiting the binding of transcription factors or by recruiting DNA binding proteins that remodel chromatin structure (Gaunt et al., 2016). For this reason, the establishment and maintenance of DNA methylation patterns are crucial for normal cellular function. Methylation patterns are established during embryogenesis in a spatiotemporal manner and are dynamic across an individual's lifetime (Hirasawa et al., 2008). The erasure of DNA methylation patterns of the gametes in the zygote occur immediately after fertilization (Trerotola et al., 2015) and are re-established as embryonic implantation occurs (Messerschmidt et al., 2014). This implicates this early stage of development as a critical window in the regulation of methylation and therefore gene expression and developmental programming (Ho et al., 2012). Any alteration to the *in utero* environment during early development can lead to permanent changes in the pattern of DNA methylation (Bollati and Baccarelli, 2010). Changes in methylation status (hyper- or hypomethylation) have been implicated in the development of certain disorders and disease (Jaffe and Irizarry, 2014) and may be a potential link between genome, environment and disease (Barfield et al., 2014, Relton et al., 2012).

1.4.2.1 Epigenetic multigenerational inheritance

One of the more fascinating and significant findings related to the DOHaD paradigm is that disease risk can be transmitted across generations. This indicates that not only are the somatic

cells perturbed but so are the gametes which will give rise to the next generation. Multigenerational inheritance is when genetic information is inherited from the F0 generation (mother) by the F1 (child) or F2 (grandchild) generation. Epigenetic inheritance is therefore the transmission of epigenetic marks from one generation to the next (Jirtle and Skinner, 2007). These epigenetic marks can be induced by environmental factors such as nutrition, temperature, stress and environmental toxicants (Hanson and Skinner, 2016). Epigenetic multigenerational inheritance occurs at a critical window of exposure linked to the development of the germ cells (sperm and egg)(Hanson and Skinner, 2016). When germ cells are developing, all epigenetic marks are erased through the process of reprogramming, allowing them to become pluripotent cells (Heard and Martienssen, 2014). Epigenetic marks are added back in a time- and sex-specific manner. Thus if germline reprogramming fails, these epigenetic marks can be retained and transmitted from one generation to the next. At fertilization, the sperm and egg provide their epigenome which has been modified during spermatogenesis and oogenesis, respectively. Even the smallest change to the epigenomes by various environmental factors can have significant effects on the developing offspring. The exposure of maturing oocytes and developing pre-implantation embryos to maternal diabetes is sufficient to re-program the foetal epigenome permanently, resulting in significant morphological changes (Hanson and Skinner, 2016).

1.4.2.2 Evidence for DNA methylation influencing early life exposure to disease risk in animal studies

There is an increasing amount of evidence that supports the role of environmentally-induced epigenetic changes in disease susceptibility. Animal studies have demonstrated that nutritional factors can modify the epigenome of the developing offspring (Lehnen et al., 2013, Brenseke et al., 2013, Plagemann et al., 2009). One of the most impressive animal model examples of the epigenetic effects of maternal nutrition on the foetus is the viable yellow agouti (A^{vy}) mouse model, in which coat colour variation is correlated to epigenetic marks established in early development (Dolinoy, 2008, Dolinoy and Jirtle, 2008, Vickers, 2011). The agouti gene is present in all mammals and functions in the determination of coat colour. Transcription of this gene occurs in the skin only for a short period, after which it is silenced

through DNA methylation (Blewitt and Whitelaw, 2013). A transposable intracisternal A particle (IAP) element was inserted upstream of the agouti gene transcription start site. This created a metastable epiallele that could be switched on or off during early development (Kanherkar et al., 2014). The degree of methylation that occurs at this IAP element correlates inversely with agouti gene expression and hence the phenotype of the mouse (Dolinoy and Jirtle, 2008, Lehnen et al., 2013). When the degree of methylation at the epiallele was increased by adding methyl donors to the mother's diet (Daxinger and Whitelaw, 2010), the offspring appeared healthy and displayed a brown coat. This brown coat colour or wild type phenotype is the result of hypermethylation at the IAP element which suppresses agouti gene expression. Hypomethylation of the IAP element increases agouti expression and results in a yellow coat phenotype which is correlated with the susceptibility to metabolic diseases, cancers and obesity (Lehnen et al., 2013).

In a study done by Plagemann et al. (2009), epigenetic changes in the hypothalamic proopiomelanocortin (*Pomc*) and insulin receptor (*Insr*) genes were associated with neonatal over feeding of rats (Plagemann et al., 2009). The level of DNA methylation was directly dependant on the amount of glucose given to the rats. Another example of environmentally-induced epigenetic changes is reduced pancreatic and duodenal homeobox 1 (*Pdx1*), also known as insulin promoter factor 1, gene expression in a rat model of intrauterine growth restriction (Brenseke et al., 2013). A reduced expression of the *Pdx1* gene (transcription factor necessary for the development and function of the insulin producing pancreatic β-cell) was observed in the β-cells of the rats experiencing intrauterine growth restriction. These rats went on to develop T2D in adulthood. A cascade of epigenetic modifications mediates the reduced *Pdx1* expression (Park et al., 2008).

1.4.2.3 Evidence from human studies

Although data from humans is still limited, the Dutch Hunger Winter famine of 1944-1945 has been used by various investigators as an equivalent to an experimental study to investigate the long-lasting effects that an nutritionally adverse *in utero* environment has on health and disease in humans (Hanson and Gluckman, 2014). Heijmans et al. (2008) showed that exposed individuals had, almost 60 years later, less DNA methylation at the imprinted insulin growth

factor 2 (*IGF2*) gene when compared to their same-sex, unexposed siblings (Heijmans et al., 2008). This finding reinforced the idea that early life exposures can cause epigenetic changes that persist throughout a person's life (Ruchat et al., 2013a, Ruchat et al., 2013b). In 2009, Tobi et al, investigated the methylation levels of 15 genes implicated in growth and metabolic disease in exposed and unexposed individuals from the Dutch Hunger Winter Cohort. They observed that the methylation of the insulin induced protein factor gene (*INSIGF*) was lower and methylation of interleukin-10 (*IL10)*, leptin gene (*LEP)*, ATP-binding cassette, sub-family A *(ABCA1)*, GNAS antisense RNA (*GNASAS)* and maternally expressed gene 3 (*MEG3)* was higher among exposed individuals when compared with their unexposed same-sex siblings. They also observed a significant interaction between the sex of the individuals and level of methylation in *INSIGF*, *LEP* and *GNASAS*. Exposed individuals had significantly altered levels of DNA methylation in a number of genes as well as a higher incidence of chronic diseases such as obesity, cardiovascular disease and diabetes when compared to unexposed siblings. The findings from these studies support the hypothesis that DNA methylation changes that occur in certain genes may be the result of exposure to prenatal famine and that the level of methylation change depends on the sex of the exposed individual (Tobi et al., 2009).

In a study of two prospective cohorts, Godfrey and colleagues measured the methylation status of CpGs in the promoters of candidate genes using DNA extracted from umbilical cord tissue obtained at birth in children who were later assessed for adiposity at 9 years of age (Brenseke et al., 2013, Godfrey et al., 2011). They found a correlation between the methylation status of the retinoid X receptor-α (*RXRA*) (in cohort 1 and 2) and endothelial nitric oxide synthase (*eNOS*) (in cohort 1 only) gene at birth with greater adiposity in later childhood. This observation suggests that epigenetics is involved in foetal programming of later obesity (Brenseke et al., 2013). In 2013, Ruchat et al, reported on the impact of GDM exposure on offspring DNA methylation levels across the genome in placental and cord blood samples (Ruchat et al., 2013a). They observed that a large number of genes in the placenta and cord blood are differentially methylated between samples from foetuses exposed or not exposed to GDM and these genes are predominantly involved in metabolic disease pathways. They also observed a correlation between the level of DNA methylation at 326 genes in the placenta and 117 genes in cord blood with neonate birth weight. In a more recent study, the

DNA methylation profiles of 14 metabolic programming candidate genes were analyzed in cord blood and placental samples (El Hajj et al., 2013). The maternally imprinted *MEST* gene and non-imprinted glucocorticoid receptor (*NR3C1*) gene showed a significant decrease in the DNA methylation levels in GDM samples when compared to controls in both tissue types (Bouchard, 2013, Ruchat et al., 2013a).

In 2012, Khulan et al, aimed to determine whether periconceptional maternal micronutrient supplementation affects genome-wide methylation within gene promoters in the foetus. They took cord blood samples from offspring of Gambian mothers who were taking micronutrient supplementations, or placebos, during the pre- and periconceptional period. They observed a significant association between micronutrient supplementation and changes in DNA methylation of CpG loci. These significant changes in the epigenome in cord blood DNA were also present in infant blood DNA samples taken at 9 months, proving that a majority of these changes are persistent. These results not only highlight the importance of micronutrient supplementation during the rainy season and around the time of conception in Gambia but also support the idea that the nutritional environment in which a foetus develops can influence the epigenetic programming of gene activity later in life (Khulan et al., 2012).

1.4.2.4 Factors causing DNA methylation variation

When interpreting the role of epigenetic variation in a complex disease, it is important to consider all the factors that may be a source of epigenetic variation (Figure 1.4). It is well known that environmental and fixed genetic factors are a source of epigenetic variation observed between individuals (Teh et al., 2014). There is also a large amount of epidemiological data that links disease risk directly to the *in utero* environment which affects the epigenome through stable epigenetic modifications. These epigenetic modifications may alter the physiology to influence disease risk in adulthood (McRae et al., 2014, Teh et al., 2014). It has been shown that DNA methylation is highly divergent between different population groups, and that this divergence may be due in large part to a combination of differences in allele frequencies and complex epistasis or gene-environment interactions (Fraser et al., 2012). DNA methylation patterns have been shown to vary between sexes depending on disease and tissue studied.

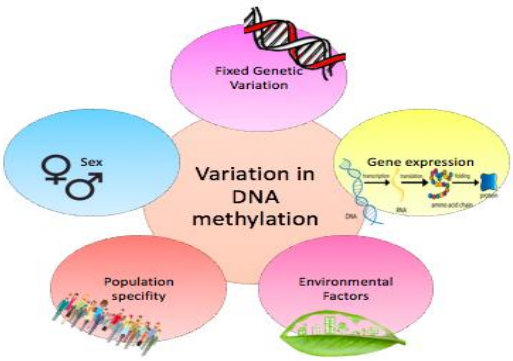


Figure 1.4: Factors influencing DNA methylation variation. The level of DNA methylation observed between individuals may vary due to a number of different factors.

1.4.2.4.1 Fixed genetic variation

Recent studies have shown that a large proportion of DNA methylation variability across individuals and populations is a result of underlying genetic variability (McRae et al., 2014, Teh et al., 2014, Wagner et al., 2014). McRae et al. (2014) found that the majority of the similarity in DNA methylation levels between relatives is due to genetic effects. This means that 20 % of DNA methylation variation that exists between individuals in a population is due to sequence-based DNA variants (SNPs) that are not located within the CpG sites. These identified SNPs, whose genotypes correlate with levels of DNA methylation, are termed methylation quantitative trait loci, or meQTLs (McRae et al., 2014). Therefore the inter-individual variation in DNA methylation is, in part, a result of nucleotide polymorphisms (Teh et al., 2014). Bell et al. (2011) utilized the Illumina HumanMethylation27 BeadChip to map associations between SNPs and methylation levels at 22,290 CpG dinucleotides in lymphoblastoid cell lines (LCLs). They found 180 CpG sites associated with nearby SNPs (Bell et al., 2011, El Hajj et al., 2014a). In a similar study using the same DNA methylation platform, Gibbs et al. (2010) studied samples from four human brain regions in 150 individuals and reported hundreds of SNP-associated CpG sites in each brain sample, with meQTLs typically located very close to the associated CpG site (Gibbs et al., 2010, Wagner et al., 2014). Although the influences of prenatal environment on future disease risk are intensively studied, it is important to address the degree to which the environmental influences are moderated by genotype. Therefore any study aiming to explore the role of DNA methylation variation to a

complex disease should carry out a parallel analysis of underlying genetic variation. These findings highlight the importance of understanding the genetic diversity of target populations.

1.4.2.4.2 Population specificity

Recent advances in high throughput technologies for measuring quantitative locus specific and genome-wide DNA methylation have provided an opportunity to characterize methylation patterns in the context of human genome variation. DNA methylation can differ in diseases and cell types, or even between monozygotic twins, but while most research has focused on variation at the cellular level, relatively little research has been done to examine how epigenetic variation affects humans at the population level. DNA methylation patterns are important for establishing cell, tissue and organism phenotypes, but very little is known about their contribution to natural human variation (Heyn et al., 2013). As the scale of DNA methylation association studies approaches that of genome-wide association studies, issues such as population stratification have to be addressed. In 2012 Fraser et al, demonstrated a wide range of within population variability in the methylation of individual CpG sites. In addition to the variation within each population, they observed that a third of the genes they studied showed differences in the DNA methylation patterns between the populations. These results suggest that DNA methylation is highly divergent between populations and this is due to a combination of genetic factors and complex gene-environment interactions. It has also been observed that distinct epigenetic and genetic signatures in certain diseases are dependent on the ethnicity of the patient (Nieminen et al., 2012). These small but extensive epigenetic differences observed between populations are most likely the result of both genetic (Barfield et al., 2014) and environmental factors (Fraser et al., 2012).

1.4.2.4.3 Environmental factors

With regard to the environment, population specific environmental factors such as socio-economic status (SES), infections and lifestyle may also contribute to differences in DNA methylation. SES is a measure of an individual's or family's economic and social position based on education, income, and occupation and has long been a strong predictor of health. In many African countries, there exist inequalities in SES and the burden of infectious and non-

communicable disease is greater among individuals of lower SES (Ataguba et al., 2011). In a preliminary study, Borghol et al. (2012) revealed an association between exposure to a low SES in childhood and differential DNA methylation in adulthood (Borghol et al., 2012). Forty adult males from the 1958 British Birth Cohort were selected according to their SES at childhood and in mid adulthood. The SES scores were determined using specific criteria, dividing the men into the most disadvantaged and the least disadvantaged. The DNA methylation profiles in adult blood indicated greater association with childhood SES than with adult SES. In a separate study, McGuinness et al. (2012), investigated the relationship between SES and DNA methylation in a subset of individuals from the Psychological, Social and Biological Determinants of Health (pSoBid) cohort. This cohort is characterized by an extreme socio-economic and health gradient. They observed global DNA hypomethylation in the individuals classified as the 'most socio-economically deprived'. They also found an association between global DNA methylation and biomarkers of cardiovascular disease (CVD) and inflammation, after adjustment for socio-economic factors. Both these studies showed that there is an association between epigenetic modifications and SES. This relationship has direct implications for population health and is reflected in further associations between global DNA methylation content and emerging biomarkers of non-communicable disease (McGuinness et al., 2012). Another example of an environmental stress factor that influences epigenetic modifications in humans is that of bacterial infections, a common cause of illness and death in SSA. Studies have shown that bacteria have the ability to change the chromatin structure and transcriptional activity of their host cells. This is achieved through epigenetic modifications such as DNA methylation. These bacterial induced epigenetic modifications may affect the host cell function either to promote host defense or to allow pathogen persistence (Bierne et al., 2012).

1.4.2.4.4 Sex

In a number of epigenetic studies, it has been shown that epigenetic modifications are largely sex-specific (Boks et al., 2009, Hannum et al., 2013). Khulan et al. (2014), observed a difference in the genes that were methylated in female and male foetuses exposed to an asthma environment (mothers had asthma). This emphasizes that different developmental

trajectories are followed by males and female under adverse intrauterine environments. The greater number of loci undergoing differential methylation in males indicated that they are impacted more than loci in females during early postnatal development. In a study done by El-Maarri et al. (2007), they measured DNA methylation in total blood in 96 healthy human males and 96 healthy human females. Global methylation was estimated by studying two repetitive DNA elements (Line-1 and Alu repeats) while single loci were investigated for three differentially methylated regions (DMRs) at *PEG3*, *NESP55* and *H19* imprinted genes and two additional loci at Xq28 (*F8* gene) and at 19q13.4 (locus between *PEG3* and *USP29*). They found that in all the studied CpGs there was slightly higher methylation in males than observed in females (El-Maarri et al., 2007). Hall et al. (2014) aimed to investigate the impact of sex on the genome-wide DNA methylation pattern in human pancreatic islets from 53 males and 34 females using the Infinium HumanMethylation450 Bead-Chip. They aimed to describe sex differences in the methylome and transcriptome in human pancreatic islets. They identified both chromosome-wide and site-specific sex differences in DNA methylation on the X-chromosome of human pancreatic islets. They also observed a higher insulin secretion in the pancreatic islets from females when compared with males (Hall et al., 2014). The observed sex differences in levels of methylation could be due to the process of X-chromosome inactivation, the presence of an additional X-chromosome in female cells or the result of downstream effects of sex determination (El-Maarri et al., 2007).

1.4.3 The role of DNA methylation in gene expression

For gene transcription to occur, the gene promoter should be readily accessible to transcription factors and other regulatory elements (e.g. enhancers). Generally, when CpG islands in gene promoter regions are methylated, transcription is repressed and the expression of the gene is silenced (Moore et al., 2013). In contrast, when promoter region CpG islands are unmethylated, transcription is initiated and gene expression is activated (Deaton and Bird, 2011). DNA methylation alters gene expression levels primarily through regulating methylation dependent interactions with transcriptional activators or repressors, and chromatin remodeling enzymes (Moore et al., 2013). Several classes of methyl-DNA binding proteins bind to methylated DNA and repress transcription, either by directly disrupting the formation of the RNA polymerase complex and associated factors at the transcriptional start

site, or by recruiting other chromatin modifiers that result in impaired transcription (Tate and Bird, 1993). These modifiers may induce changes in post-translational histone modification states leading to the formation of inactive chromatin and the prevention in the formation of the RNA polymerase complex (Conerly and Grady, 2010).

Gene expression is a heritable trait and its variation is one of the main driving mechanisms underlying complex disease susceptibility (Cookson et al., 2009). The sequencing of the human genome has allowed for the detection of disease-causing mutations in many Mendelian disorders as well as the identification of significant associations between polymorphisms and complex disease (Costa et al., 2013). However, identifying causative variations for multifactorial diseases, such as diabetes, remains a complex task (Twine et al., 2011). Gene expression analysis provides a valuable understanding of the normal biological and disease processes: alterations in gene expression patterns are often responsible for the differences observed between disease and healthy states (Shaat and Groop, 2007). Although the genome gives a static view of the genetic and regulatory information defining an organisms phenotype, knowledge of an organism's transcriptome (the entirety of transcribed genes) is essential (Kratz and Carninci, 2014). Whole-transcriptome analysis represents a powerful tool for providing not only insight into the functional elements that contribute to the current genetic knowledge of diseases (Costa et al., 2013), but also information that is necessary for the complete understanding on how the same genome can produce different cell types in an organism and how these genes are regulated in health and disease (Kratz and Carninci, 2014).

There are two main approaches used to study gene expression, namely a genome-wide or a targeted approach. The method chosen depends on the information one has available. When the key genes of interest are not known, genome-wide approaches such as the older sequence-based microarray technology or the more modern high-throughput sequencing-based RNA-Seq technology are used (Mortazavi et al., 2008). These methods revolutionized expression profiling by enabling the measurement of thousands of genes simultaneously. For many years, hybridisation-based microarrays were the dominant platform for high throughput analysis of gene expression (Zhao et al., 2014) but they have several limitations. These include the need for a priori knowledge of sequences to analyse, an inability to detect splice site

isoforms and novel genes/exons (Costa et al., 2013), background noise due to cross-hybridization, and poor detection and quantification of low expressing transcripts (Kratz and Carninci, 2014). The use of RNA-seq is fast becoming a more affordable and sensitive competitor for differential expression analysis (Frazee et al., 2014).

1.4.4 Limitations to epigenetic studies

DNA methylation patterns are cell type and tissue specific (Barouki et al., 2012, Lehnen et al., 2013). For this reason, it is important to assess epigenetic modification in tissues that contribute to the disease being studied. However, in epidemiological studies conducted in humans, obtaining relevant tissue is often invasive and/or impossible (Jaffe and Irizarry, 2014, Petronis, 2010). For this reason, surrogate tissues, most commonly whole blood, are used (Paul and Beck, 2014). The cell types present in the surrogate tissue will reflect epigenetic modifications found in the target tissue, or at least yield biomarkers that, although not directly causative of the disease, can still be used for predictive and/or diagnostic purposes as well as provide new pathophysiological insights into the disease (Lowe and Rakyan, 2014, Rakyan et al., 2011). The cause-and-effect relationship between disease and epigenetic modifications is complex. The link between epigenetic modification variation observed between affected and unaffected individuals and disease (or predisposition to developing the disease) is difficult to determine as these epigenetic changes may simply reflect the differences in cellular composition between the disease and non-disease state (Lowe and Rakyan, 2014, Petronis, 2010, Verma, 2012). The disease itself could induce epigenetic changes so it is important to distinguish between causal and non-causal associations (Petronis, 2010). It is important to account for cellular heterogeneity in whole blood. This can be done by using a *post hoc* bioinformatics solution for confounding cell-type bias in EWASs (Jaffe and Irizarry, 2014, Zou et al., 2014). Another challenge facing EWAS studies is determining the optimal sample size. It is assumed that if the anticipated effects are small, thousands of participants should be included to avoid type one errors (Paul and Beck, 2014). It is important to perform a power analysis, which is used to determine the minimum sample size for a study to give statistically significant results (Petronis, 2010). Careful phenotyping can improve statistical power especially in studies with small to moderate sample sizes (Sham and Purcell, 2014).

Another important limitation to consider when designing epigenetic and gene expression experiments is the effect of confounding factors. For example, a possible confounding factor is that environmental stress can alter the relative abundance of different cell types in the blood which could lead to altered levels of measured DNA methylation for specific genes (Feil and Fraga, 2012). DNA methylation is associated with cell fate determination in haematopoiesis, and its perturbation could affect the cell populations that constitute peripheral blood (Borgel et al., 2010). To avoid such confounding effects, and if experimental design allows, specific cell lineages and/or tissues should be studied (Feil and Fraga, 2012). There are a number of methods that have been proposed to minimize the effects of confounding. In studies of the association of biomarkers with disease, the technique of Mendelian randomization (Davey Smith and Ebrahim, 2004) has received much attention as a way of overcoming reverse causation and uncontrolled confounding. They use a gene as an instrumental variable to assess the causal effect of the biomarker on disease risk. Relton and Davey Smith (2012) have proposed a novel two-step extension of this idea for methylation studies, using two genes as instrumental variables, one to estimate the exposure–methylation association, the other to estimate the methylation–disease association (Relton and Davey Smith, 2012).

1.5 RNA-sequencing (RNA-seq): a revolutionary tool for transcriptomics

RNA-Seq is a widely used and powerful next generation sequencing technique to comprehensively study the entire transcriptome (Sultan et al., 2012) and allow one to investigate the expression levels and structure of transcripts without prior knowledge of the transcriptome content (Costa et al., 2010). This enables the discovery of novel transcripts and also provides the potential to reveal novel molecular biomarkers for human diseases, through comparison of the transcriptome from normal and diseased samples (Kratz and Carninci, 2014).

RNA-seq has emerged as a powerful tool in the study of complex human diseases. A number of genes have been identified and implicated in the development of T2D by transcriptome

analyses in human or animal models (Ghosh et al., 2010), however, in comparison to T1D and T2D, GDM has been subject to fewer transcriptome analyses (Evangelista et al., 2014). Nonetheless, many genes related to obesity and oxidative stress have been shown to be associated with both T2D and GDM (Watanabe et al., 2011). Transcriptome signatures obtained from placenta and whole blood cells have identified genes involved with lipid metabolism that are differentially expressed between T2D and GDM (Zhao et al., 2011; Radaelli et al., 2009). Recently in a meta -analysis of the transcription profiles of T1D, T2D and GDM patients, Evangelista et al. (2014) found that the gene expression profiles of GDM patients were more closely related to T1D patients than T2D patients. However, the analysis of these gene expression signatures was impaired by the presence of multiple variables associated with each type of diabetes (Evangelsita et al., 2014). A GDM transcriptome analysis performed in 2014 by Donadi et al. (2014) demonstrated an increased expression of a number of genes related to the major histocompatibility complex, namely *HLA-DRB6, DQB1, DQB2, DOA* and *DQA2* (Donadi et al., 2014). The modulation of these transcripts in GDM patients reinforces the hypothesis of a deregulation of HLA class II genes in GDM patients. Twine et al. (2011) provided an extensive transcriptome analysis of post-mortem frontal and temporal lobes of patients with Alzheimer's disease, highlighting a differential expression of known causative genes and also of previously unannotated expressed regions (Costa et al., 2013, Twine et al., 2011). The introduction of RNA-seq into cancer research has had a large and positive impact on this area of research. Using this method to investigate cancer transcriptomes may provide answers to the multitude of questions about carcinogenesis in humans (Costa et al., 2013). Many RNA-seq studies have implicated alternative splicing and detrimental fusion transcripts in the carcinogenesis of different tissues and organs (Edgren et al., 2011, Hong et al., 2016, Nacu et al., 2011). RNA-seq was used to identify genes differentially expressed between individuals with heart failure (N=6) and those with non-failing hearts (N=6). Using the genes identified from this small dataset, they were able to accurately classify heart failure status in a larger cohort (N=313). Their results indicate that, using a small training dataset, it is possible to use RNA-seq to classify disease status for complex diseases (Liu et al., 2015).

1.5.1 Challenges for RNA sequencing

Although RNA-seq has brought a significant qualitative and quantitative improvement to transcriptome analysis (Labaj et al., 2011, Shendure, 2008, Shendure and Ji, 2008), there are limitations from sample preparation to data analysis. Many procedures are involved in the preparation of samples for RNA-seq, namely extraction, fragmentation, reverse transcription and amplification, all of which are susceptible to experimental bias (Costa et al., 2013). Unlike small RNAs, which can be sequenced directly after adapter ligation, large RNA molecules must be fragmented into smaller pieces to be compatible with deep-sequencing technologies (Wang et al., 2008). Fragmentation is the process that involves the breakdown of large RNA molecules into smaller fragments and the advantage of this process is that it reduces the formation of secondary structures (which reduce the ability of RNA to be fragmented), allowing higher sequence coverage across the length of the transcript (Costa et al., 2013). The presence of 'susceptibility fragmentation sites' can alter the representation of that sequence within the library (Sendler et al., 2011). The probability of random fragmentation can be altered by the GC content of the transcripts. This affects the counting efficiency and will present a severe bias in gene expression measurements (Costa et al., 2013). The GC content has also been shown to affect the cDNA amplification process because GC rich fragments tend to form double stranded or highly paired secondary structures which affect the action of reverse transcription (Aird et al., 2011). During the process of reverse transcription, the cDNA strand that is being synthesized could dissociate from the template RNA and re-anneal to an alternate piece of RNA of similar sequence. This process of 'template shifting' generates artificial chimeric cDNAs (Ozsolak and Milos, 2011) and can inhibit the identification of exon-intron boundaries and true chimeric transcripts (Cocquet et al., 2006, Roy and Irimia, 2008). Issues also arise with the reverse transcriptase enzymes, as they tend to have a low fidelity and variable RNA to cDNA conversion efficiency when compared to other polymerases (Ozsolak and Milos, 2011). Another important issue to consider is the coverage versus cost, Greater coverage (percentage of transcripts measured) requires more sequencing depth, which comes at a larger cost, Considerable sequencing depth is required to detect rare transcripts (Wang et al., 2009). Not only does RNA-seq face many experimental challenges, but like many high throughput sequencing technologies, this method also faces bioinformatics

challenges. These include the development of methods that can efficiently store, retrieve and process large amounts of data, remove low quality reads and be able to reduce errors in image analysis (Wang et al., 2009).

Motivated by the advantages of RNA-Seq technology for gene expression profiling, and given the clinical and public health significance of GDM, we sequenced the transcriptomes of six women who developed the disease and 6 healthy controls in order to potentially identify genes that are associated with the development of gestational diabetes. We also sequenced the transcriptomes of placental tissue that was exposed to an adverse *in utero* environment (presence of GDM) and unexposed placental tissue, to see if exposure had an effect on placental gene expression. To date, no genome-wide gene expression RNA-seq studies for GDM have been done in black South Africans.

1.6 Specific Aim and Objectives

The main aim of this research study was to determine whether genes involved in metabolic processes had differential expression due to promoter region methylation and whether this alteration occurred as an adaptive response to exposure to an adverse *in utero* environment inflicted by the presence of GDM.

Objectives (Figure 1.5)

1. To assess differential genome-wide gene expression in blood samples from women with GDM and controls to identify genes that display statistically (qvalue<0.05) significant differential expression (≥2 fold).

2. To assess differential genome-wide gene expression in placental tissue from female neonates born to women with GDM (exposed) and those born to controls (not exposed; absence of GDM) to identify genes that display statistically (qvalue<0.05) significant differential expression (≥2 fold).

3. To identify genes that are significantly (qvalue<0.05) differentially expressed (≥2 fold) in

both blood and placental tissue and to determine whether their expression pattern is due to aberrant promoter region DNA methylation.

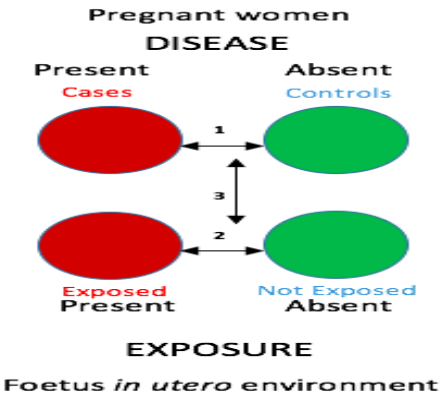


Figure 1.5: **The different comparisons that will be made in this research study**

CHAPTER TWO

2. MATERIALS AND METHODS

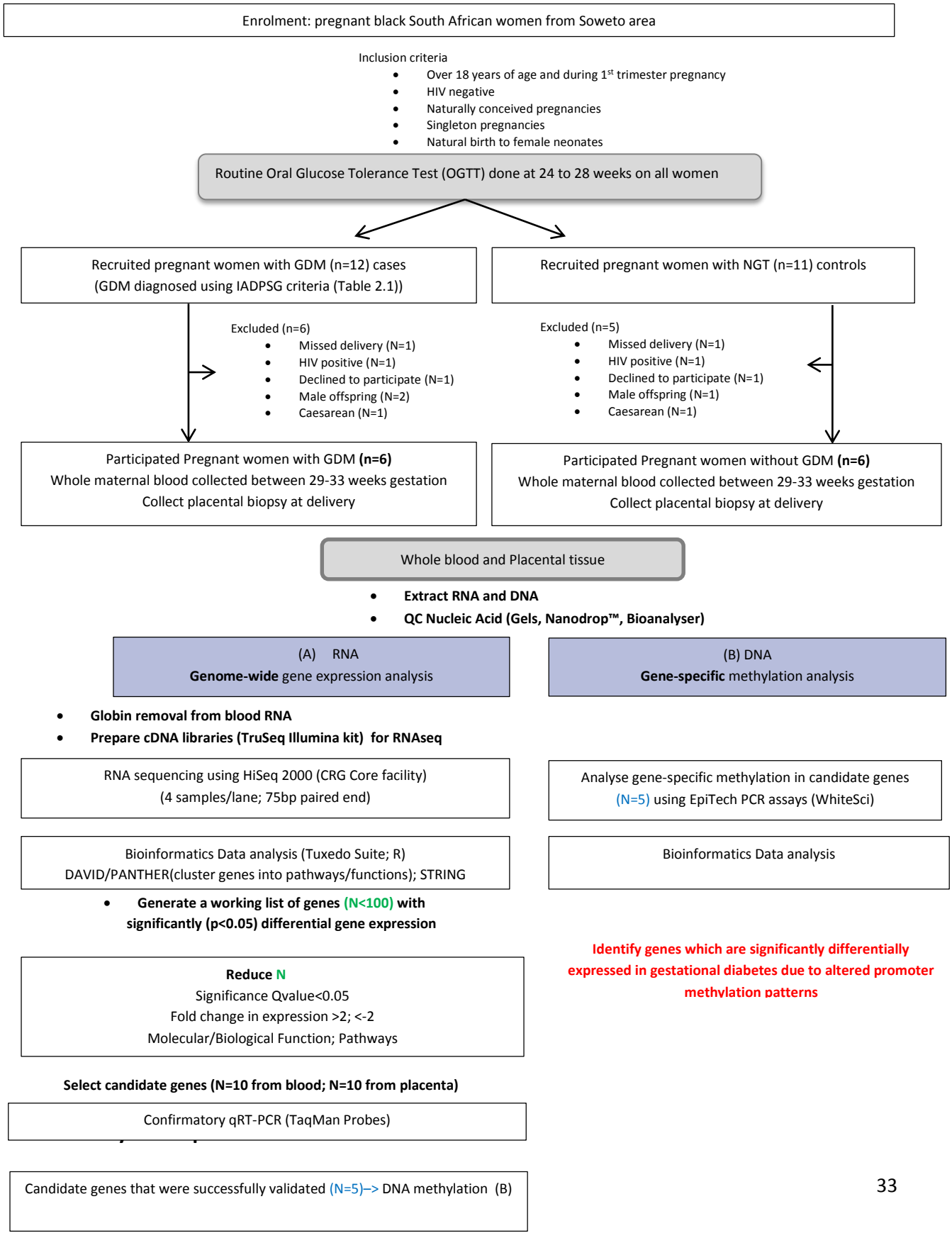


Figure 2.1: An overview of the materials and methods used in the study

2.1 Study Participants

Participant recruitment and sample collection was done in collaboration with the Developmental Pathways and Health Research Unit (DPHRU) at Wits, under the umbrella of the Foetal/Soweto baby growth study. The Foetal/Soweto baby growth study aims to recruit over 3000 participants. It is from this larger cohort that participants for this study were recruited. The target population is black South African women in their first trimester of pregnancy who present at clinics around the Soweto, Johannesburg area. All women recruited were followed from 1st trimester of pregnancy (<14 weeks) to delivery, with five to six visits depending on gestational age at entry into the study (visit 1: <14 weeks; visit 2: 14-18 weeks, visit 3: 19-23 weeks; visit 4: 24–28 weeks; visit 5: 29–33 weeks; visit 6: 34–38 weeks). At each visit, anthropometric variables such as blood pressure and weight were measured using standardized procedures. Glucose tolerance was assessed using a 75 g Oral Glucose Tolerance Test (OGTT) performed at approximately 24 to 28 weeks gestation. The new International Association of the Diabetes and Pregnancy Study Groups criteria (IADPSG) were used to diagnose GDM (Table 2.1).

Table 2.1: **IADPSG diagnostic criteria for GDM** (Duran et al., 2014)

	Venous plasma glucose threshold (mmol/L)
Fasting	**≥ 5.1**
75 g OGTT: 1 hour	**≥ 10.0**
75 g OGTT: 2 hours	**≥ 8.5**
One or more values equal or exceeding diagnostic threshold is indicative of gestational diabetes mellitus	

Women with one or more OGTT values equal to or exceeding the above mentioned diagnostic thresholds were selected as cases for the study. Women with OGTT results lower than the specified diagnostic threshold at each time point were selected as controls. These selected women were asked to provide informed consent (Appendix Ai) and were given an information sheet (Appendix Aii) that they were asked to sign if they understood the information and

agreed to participate before their inclusion in the study. The University of the Witwatersrand Human Research Ethics Committee (Medical) approved the project (Ethics clearance No: M130420)(Appendix Bi and Bii).

Initially 23 women were recruited to participate in the study, however, following the strict exclusion criteria, 11 women were excluded from the study (Figure 2.1). At the time of delivery, seven women were excluded from the study. Three gave birth to male offspring, two gave birth at a different hospital and it was not possible to obtain the placental sample and there were two missed deliveries (the women failed to contact the nurses prior to delivery). The total number of participants for which both blood and placental tissue was obtained was 12 (6 cases and 6 controls). Studies have shown that there are sex-specific differences in placental global gene expression (Osei-Kumah et al., 2011). For this reason, same sex offspring (only females) will be used in this study. Venous whole blood was obtained from the women at their 5th visit once the OGTT test results were available. Placental samples were taken within an hour of birth. Control individuals had normal glucose tolerance (NGT) during pregnancy and were as closely matched as possible by age and BMI to case subjects (women with GDM).

2.2 Sample collection

2.2.1 Whole Blood Samples

Two whole blood samples were taken from each of the women at their 5th visit (29-33 weeks gestation). A 6 ml blood sample was collected in a purple top EDTA tube (for DNA extraction). These tubes were stored at 4°C until needed. The second blood sample (3 ml) was collected in a blue top Tempus™ tube (for RNA extraction). Two separate blood collection tubes were necessary because of the different stabilization and storage conditions for RNA and DNA. The Tempus™ tubes contain a RNA Stabilizing Agent that, once mixed with whole blood, immediately begins to lyse the cells. The stabilizing reagent then inactivates cellular RNases and selectively precipitates RNA. After the blood was drawn into the Tempus tubes, they were vortexed vigorously for 10 seconds to ensure proper mixing of the blood with the stabilizing reagent. These tubes were stored at -80°C until needed.

2.2.2 Placental Tissue Samples

At the time of delivery, the placenta was obtained and two 8 mm vertical placental punch biopsies were taken within an hour of birth. These biopsies were taken from the placental disc, avoiding the umbilical cord insertion site and approximately 3 cm from the edge of the placenta (Figure 2.2). The direction of the punch biopsy was from maternal to foetal side. One biopsy (used for RNA extraction) was stored in 4 ml of RNALater solution and stored at -80°C. The second biopsy (used for DNA extraction) was flash frozen by placing the bottom of the uncapped tube (containing the placental tissue) in liquid nitrogen and also stored at -80°C until needed.

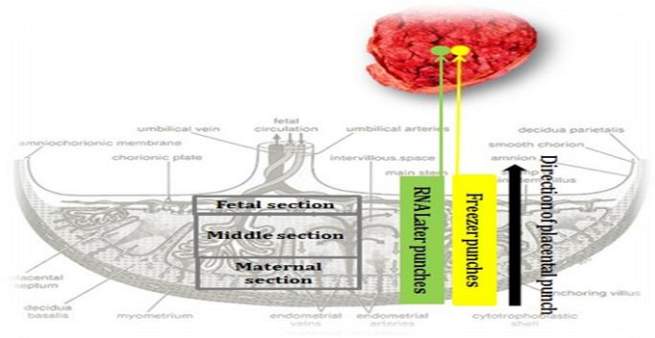


Figure 2.2: Schematic of placenta tissue sampling. Two placental biopsies were taken, the direction of the placental punches was from the maternal side through to the foetal side.

2.3 Nucleic acid (RNA and DNA) extraction

RNA (transcribed into cDNA) was used for the gene expression analysis and DNA for methylation analysis. RNA extraction and expression validation by qRT-PCR, as well as methylation studies were performed at Wits. The mRNA library construction and RNA-seq were performed at the Centre for Genomic Regulation (CRG), Barcelona, Spain. Data analysis was started in Spain and concluded at Wits.

2.3.1 From whole blood

Total RNA was extracted from whole blood following the Tempus™ Spin RNA Isolation Kit Protocol (Applied Biosystems). The reagents and consumables included in this kit allow the isolation of 6 to 25 µg of high quality RNA from 3 ml of whole blood)(Table 2.2). Before

extraction, the frozen stabilized blood was thawed to room temperature and transferred to a 50 ml tube containing 1X calcium- and magnesium-free phosphate-buffered saline (PBS). After a centrifugation step, the RNA-containing pellet was resuspended and transferred onto a purification filter. The purified RNA was eluted following a micro-centrifugation step (For a full description of the extraction protocol, refer to Appendix Ci).

DNA was extracted from whole blood samples using the salting out method (Miller et al., 1988). This two-day method of extraction yields a high quantity of DNA (ranging from 200 – 1000 ng/µl)(Table 2.2). Nuclear lysis buffer was added to the blood to break down the red blood cells. The cell lysates were digested overnight with a SDS and Proteinase K solution. Saturated NaCl was added to the mixture and centrifuged. This step precipitates the protein which was then removed. Ethanol was added and the tubes inverted until the DNA precipitate formed. The DNA was removed with a pipette tip and transferred to a microcentrifuge tube containing 200 µl of TE buffer. The resuspended DNA was stored at 4°C until needed (Full protocol in Appendix Cii).

2.3.2 From placental tissue

Total RNA was extracted from the placental tissue using the RNeasy Mini Kit (Qiagen). This kit provides fast purification of high-quality RNA from tissues using silica-membrane RNeasy spin columns with a binding capacity of 100 µg RNA. The maximum weight of tissue that can be processed for each sample with this kit is 30 mg (Table 2.2). After thawing the placental biopsies at room temperature, a 30 mg sample was excised (from just behind the membrane, ensuring that the placental sample was largely foetal) from each of the 8 mm placental biopsies and placed in a tube containing mercaptoethanol and efficiently disrupted using a TissueRuptor system (Qiagen). After the samples had been lysed and homogenized, ethanol was added to the lysate. The lysate was loaded onto the RNeasy silica membrane. Up to 100 µg of RNA binds to the membrane and the contaminants are efficiently washed away. Pure RNA is eluted in 50 µl of nuclease free water (For a full description of the extraction protocol, refer to Appendix Ciii).

DNA was extracted from placental tissue using the DNeasy Blood and Tissue kit (Qiagen). Less than 25 mg of placental tissue (Table 2.2) was cut into small pieces and placed into a microcentrifuge tube containing Buffer ATL and Proteinase K solution. The mixture was incubated overnight at 56°C until the tissue was completely lysed. This mixture, together with Buffer AL and ethanol was placed into a DNeasy Mini spin column and centrifuged. The lysate was then washed in two separate steps using washing buffer. The DNA was eluted from the microcentrifuge tube using an elution buffer (For a full description of the extraction protocol, refer to Appendix Civ).

Table 2.2: Extraction kits used to obtain RNA and DNA from whole blood and placental tissue

RNA Extraction				
Tissue	**Collected per woman**	**Kit**	**Input**	**Output**
Whole Blood	1x Tempus Tube	Tempus Spin RNA Isolation Kit	3 ml Blood	6-25 µg RNA
Placenta	1x 8 mm biopsy RNALater	RNeasy Mini kit	30 mg tissue	30 µg RNA
DNA Extraction				
Whole Blood	1x EDTA purple top tube	Salting Out	6 ml Blood	+/- 200 µg DNA
Placenta	1x flash frozen 8 mm biopsy	DNeasy Blood and Tissue Kit	25 mg tissue	15-30 µg DNA

2.3.3 Nucleic acid Quality and Quantity control

The total RNA and DNA extracted from whole blood and placental tissue was run on a 0.8 % agarose gel to check quality. 1 µl of each nucleic acid sample was run on the Nanodrop Spectrophotometer in order to determine quantity (ng/µl). The purity of the nucleic acids was assessed through the ratio of absorbance at 260 nm and 280 nm. Generally, a ratio of ~1.8 is accepted as "pure" for DNA and ~2.0 for "pure" RNA. If the ratio is appreciably lower in either case, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm. The DNA was stored at -20°C until needed later in study for the methylation analysis.

For RNA, the Bioanalyser (Agilent Technologies) was used as a second method to measure quality. The Bioanalyser measures the RNA Integrity number (RIN) as well as the rRNA ratio (28S/18S). The RIN score is generated for each sample on a scale from 1-10 (where 10 indicates high quality RNA). The RIN should generally be above 7 and any samples with RIN <5 should not be considered for RNA sequencing. The RIN value is not the only index of RNA integrity. Some papers indicate that the most accurate factor to evaluate the integrity of the RNA is the 28S/18S ratio which must generally be above 1.0 (Miller et al., 2004). An aliquot of the RNA samples was normalized to 400 ng/µl using nuclease free water and 1 µl of this normalised sample was loaded onto the BioAnalyser chip.

2.3.4 Globin mRNA removal from whole blood

Molecular profiles of circulating blood can be associated with physiological and pathological events occurring in other tissues and organs of the bodies (Shin et al., 2014). Peripheral whole blood is therefore a highly desirable tissue due to its accessibility and its relatively non-invasive mode of collection. Gene expression profile studies of human blood samples are, however, confronted by numerous challenges. Accounting for approximately 80 - 90 % of transcript species, globin dominates the peripheral whole blood transcriptome, potentially affecting the ability to detect other transcripts, particularly those with lower expression. The process of removing globin transcripts from RNA prior to accessing mRNA expression, is not uncommon. Shin et al. (2014) showed that globin depletion results in a statistically significant increase in the number of detectable transcripts, particularly lower expressing transcripts. Although this process significantly lowers the quality and quantity of total RNA (Shin et al., 2014; Liu et al., 2006), it does not negatively affect the relative abundance of other transcripts when proceeding to cDNA library preparation. They showed that there was no statistically significant difference in the total number of reads, total number of mapped reads or percentage of reads filtered out when comparing globin depleted RNA to non-globin depleted RNA. Overall, globin depletion appears to meaningfully improve the quality of peripheral whole blood RNA-seq data, increasing the number of detectable transcripts. A key concern with this process is the reduction in the quantity of extracted RNA especially in the cases where biological samples are rare and difficult to replace (Shin et al., 2014).

Before the preparation of the cDNA libraries, globin mRNA was removed from the total RNA extracted from the whole blood samples. This was done using the GLOBINclear™ Kit (human), for globin mRNA depletion (ThermoFischer Scientific). Briefly, this kit removes globin mRNA from total RNA via hybridization with biotinylated DNA oligos that specifically capture globin mRNA followed by binding with streptavidin magnetic beads. This removal method uses novel, non-enzymatic technology that rapidly depletes >95 % of the alpha and beta globin mRNA from total RNA preparations derived from whole blood. According to the manufacturers, this allows detection of up to 50 % more, previously undetected, genes (For the full mRNA removal procedure, see Appendix D).

2.4. A standard RNA sequencing workflow

The RNA-seq workflow, from sample preparation through to data analysis, enables rapid profiling and deep investigation of the transcriptome. RNA-seq is the simultaneous execution of millions of sequencing reactions of relatively short read length (30 – 500 bp) in parallel, generating massive amounts of sequence data per run (Shendure and Ji, 2008). The term RNA-seq denotes an ever expanding menagerie of protocols, nevertheless, they all have similar concepts: extracting cellular RNA, removing rRNAs, isolating the poly-A mRNA transcripts and converting this population of mRNA to a library of cDNA fragments, which are then sequenced (Kratz and Carninci, 2014). RNA-seq experiments must be analysed with robust, efficient and statistically principled algorithms (Trapnell et al., 2013, Trapnell et al., 2012). The direct product of an RNA-seq experiment is a large electronic file that contains millions of sequencing reads from each sample. The first step is to align the sequencing reads to the reference genome in order to know where they have originated from. Because of the massive amount of reads produced, specialized algorithms need to be used to do the alignment. These algorithms significantly increase the alignment speed by indexing the reference sequence in a way which makes it possible to quickly match the reads against the reference (Langmead and Salzberg, 2012). Following RNA-seq reads mapping, the data needs to be converted into a quantitative measure of gene expression. Because the number of reads produced from an

RNA transcript is the function of that transcripts abundance, read density can be used to measure transcript and gene expression (Cloonan and Grimmond, 2008). There are many different RNA-seq analysis packages that can be used for RNA-seq data analysis however, the Tophat and Cufflinks protocol was used in this study.

2.4.1 cDNA library preparation

The cDNA libraries were prepared using the TruSeq Stranded mRNA preparation kit (Low-Throughput protocol; Illumina) according to manufacturer's instructions. Briefly, 2 μg of total RNA sample from globin mRNA-depleted whole blood and placental tissue was used for polyA mRNA selection using polyT oligo attached magnetic beads and two rounds of purification. During the second elution of the polyA RNA, the RNA is fragmented and primed for cDNA synthesis. cDNA was synthesized from the enriched and fragmented RNA using reverse transcriptase, SuperScript II and random primers. The cDNA was converted into double stranded DNA (dsDNA) which was used for library preparation. The overhangs on the dsDNA resulting from fragmentation are then converted into blunt ends. A single 'A' nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a complementary extension for ligating the adapter to the fragment. A multiple indexing adapter is then ligated to the ends of the dsDNA, preparing them for hybridization onto a flow cell in the HiSeq 2000 (Illumina). PCR is then used to selectively enrich the DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library (For the full protocol description, see Appendix E).

2.4.1.1 cDNA library QC

The quality and quantity of the sample libraries were assessed before the sequencing procedure. To achieve the highest quality data on Illumina sequencing platforms, it is important to create optimum cluster densities across every lane of the flow cell. Optimizing cluster densities requires accurate quantitation of cDNA library templates. The concentrations of the libraries were quantified using the Nanodrop and the size and purity (quality) were

measured using the Agilent Bioanalyzer. 1 µl of the cDNA library was loaded on the Bioanalyzer using a DNA-specific chip, the Agilent DNA 1000.

2.4.1.2 Normalization and Pooling of cDNA Libraries

RNA-seq protocols use a RNA fragmentation approach prior to sequencing to gain sequence coverage of the whole transcript. This means that long transcripts will have more reads mapping to them when compared with short transcripts of similar expression level. For this reason, read counts need to be properly normalized to extract meaningful expression estimates (Mortazavi et al., 2008). Moreover, each sequencing run has a given variability which will influence the number of fragments mapped across samples. Hence, it is also necessary to normalize for each sequencing run in order to avoid the possibility that genes will appear to be differentially expressed only as a result of the presence of more sequences in one condition when compared to another. One of the ways in which sequencing data is normalized is to use the reads per kilobase of transcript per million mapped (RPKM) metric, which normalizes a transcript's read by both its length and the total number of reads mapped in the sample. In a similar way, fragments per kilobase of transcript per million mapped (FPKM) metric normalizes paired-end data (Oshlack et al., 2010).

10 µl of the indexed cDNA libraries were normalized to 10 nM in the DCT (Diluted Cluster Template) plate using Tris-HCl 10 mM, pH 8.5 with 0.1% Tween 20, and then pooled in equal volumes in the PDP (Pooled DCT Plate). Each normalized sample library to be pooled together was transferred from the DCT plate to one well of the PDP plate. The 24 indexed cDNA libraries were pooled together in equal concentrations into 6 pools (4 samples per pool as follows: placenta_case; placenta_control; blood_case; blood_control) (Table 2.3).

Table 2.3: Sample arrangement in the flow cell for sequencing. Each pool indicates the samples run in a single lane in the HiSeq2000 sequencer. The indexes used in each pool are listed as this is how the samples are recognized. Indexes used in each lane need to be compatible**.**

Sample ID	Tissue	Case/Control	Index	Index Sequence
Pool 1				
1048b	Blood	Case	2	C G A T G T
1107b	Blood	Control	7	C A G A T C
1087p	Placenta	Control	14	A G T T C C
1054p	Placenta	Case	4	T G A C C A
Pool 2				
1067p	Placenta	Control	5	A C A G T G
1054b	Blood	Case	18	G T C C G C
1094b	Blood	Control	15	A T G T C A
1060p	Placenta	Case	16	C C G T C C
Pool 3				
1048p	Placenta	Case	5	A C A G T G
1094p	Placenta	Control	15	A T G T C A
1086b	Blood	Case	12	C T T G T A
1067b	Blood	Control	19	G T G A A A
Pool 4				
10225b	Blood	Case	2	C G A T G T
1090b	Blood	Control	4	T G A C C A
1090p	Placenta	Control	7	C A G A T C
1086p	Placenta	Case	16	C C G T C C
Pool 5				
1107p	Placenta	Control	12	C T T G T A
1060b	Blood	Case	6	G C C A A T
1087b	Blood	Control	13	A G T C A A
10276p	Placenta	Case	14	A G T T C C
Pool 6				
1061p	Placenta	Control	6	G C C A A T
10276b	Blood	Case	13	A G T C A A
10225p	Placenta	Case	18	G T C C G C
1061b	Blood	Control	19	G T G A A A

2.4.2 Sequencing the cDNA libraries

Each of the six pools (containing 4 samples each) was sequenced (75 bp, paired-end) in a separate lane on the HiSeq 2000 platform in the same sequencing run for side-by-side comparison. This sequencing depth should generate ~ 50,000,000 reads per sample. Before

analysing the sequences generated and extracting biological conclusions from them, it is critical to evaluate the quality of the sequences as well as the overall sequencing performance. Therefore, before aligning the sequencing reads to a reference genome, the low-quality bases must be removed. Quality Control (QC) of the sequences takes into account duplication rate, rRNA abundance, strand specificity, coverage continuity at all annotated transcripts and performance at 5' and 3' ends. The resulting Phred score is used to evaluate the quality of the sequencing; the content of bases; the amount of N (specific nucleotide not called) bases and the sequenced read lengths. Based on this type of analysis, the bases with low sequencing quality should be trimmed ensuring the high quality of the sequencing data. In this study, the program FASTQC was used to check the quality of high throughput sequence. FASTQC produces several quality control plots which are important when evaluating the condition of the millions of generated raw sequence files before doing any further analysis. If the quality is not optimal, trimming and filtering of sequence reads must be done, otherwise the downstream analysis will not provide statistically relevant results (Niiranen, 2015). The FASTQC files for each sample were analysed in order to determine the quality of the millions of generated sequences.

2.5. Differential gene expression analysis of RNA-seq data using tophat and cufflinks

The downstream analysis of the sequencing reads generated for each sample was done following the pipeline shown in Figure 2.3, using programs called tophat, cufflinks (cuffmerge, cuffdiff) and cummeRbund (Trapnell et al., 2012). Tophat and cufflinks are free open source software tools used for gene discovery and expression analysis of high throughput RNA-seq data. Together they allow for the identification of novel genes and splice variants as well as for the comparison of gene expression between disease and healthy states (Trapnell et al., 2012).

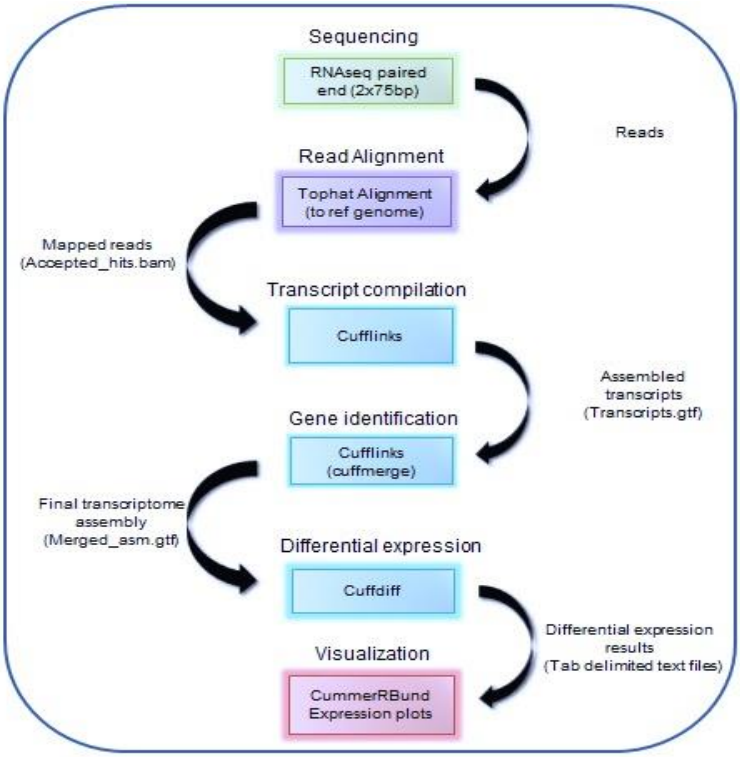


Figure 2.3: **An overview of the tophat/cufflinks RNA-seq analysis protocol.**

The RNA-seq data was aligned to the Human Reference Genome (hg19) with tophat 0.5.9-r16 (http://tophat.cbcb.umd.edu) with default options. Tophat is a fast splice site junction mapper for RNA-seq reads. The script written for tophat to align the generated reads to the genome using the ultra-high throughput short read aligner, ”Bowtie”, and then analyse the mapping results to identify splice sites between exons, is shown in Figure 2.4. Tophat generates an output file named “*accepted_hits.bam”* file. This contains all the aligned reads and was used as the input file for cufflinks. After running tophat, the resulting alignment files were provided to cufflinks (http://cufflinks.cbcb.umd.edu) which assembled the individual transcripts from the aligned RNA-seq reads, estimates their abundances, and tested for differential expression. Cufflinks produces an assembled transcriptome fragment for each sample using the “*accepted_hits.bam”* file as the input. The script written to run cufflinks is shown in Figure 2.5. The resulting transcriptome fragments of each sample are then merged together with the reference transcriptome annotation into one file for further analysis. This was done using cuffmerge. The script written to merge the cufflinks files is shown in Figure 2.6.

```
Tophat.26.sh
#!/bin/bash
 # OGE parameters
 #$ -q xe-el6
#$ -N RNAseqANGELA
#$ -e /no_backup/xe/ahobbs/alignment/0026/e26logs
#$ -o /no_backup/xe/ahobbs/alignment/0026/o26logs
 #$ -m abe
#$ -M angela.hobbs@crg.eu
 #$ -pe smp 4
#$ -l h_rt=20:00:00
#$ -l virtual_free=20G

# paths
PATH=/users/GD/tools/bowtie/bowtie2-2,1,0:$PATH
export PATH

/software/bi/el6,3/current/tophat/tophat2 --output-dir /no_backup/xe/ahobbs/alignment/0026 --num-threads 4 --rg-id 0026 --rg-library 0026 --rg-
sample 0026 --rg-platform illumina --transcriptome-index /no_backup/xe/ahobbs/alignment/index
/db/igenomes/Homo_sapiens/Ensembl/GRCh37/Sequence/Bowtie2Index/genome /no_backup/xe/ahobbs/samples/26_9966_GTCCGC_read1.fastq.gz
/no_backup/xe/ahobbs/samples/26_9966_GTCCGC_read2.fastq.gz
```

Figure 2.4: **The script written to align the generated sequences to the reference human genome using Tophat**.

```
#!/bin/bash
 # OGE parameters
 #$ -q xe-el6
#$ -N RNAseqANGELA
#$ -e /no_backup/xe/ahobbs/alignment/0009/e,cl9,logs
#$ -o /no_backup/xe/ahobbs/alignment/0009/o,cl9,logs
#$ -V
 #$ -m abe
#$ -M angela,hobbs@crg,eu
 #$ -t 1

#$ -pe smp 8
#$ -l h_rt=20:00:00
#$ -l virtual_free=40G

/users/GD/tools/cufflinks/cufflinks-2,2,1,Linux_x86_64/cufflinks --output-dir /no_backup/xe/ahobbs/alignment/0009 --num-threads 8 --max-bundle-
frags 100000000 --GTF /db/igenomes/Homo_sapiens/Ensembl/GRCh37/Annotation/Genes/genes,gtf --GTF-guide
/db/igenomes/Homo_sapiens/Ensembl/GRCh37/Annotation/Genes/genes,gtf /no_backup/xe/ahobbs/alignment/0009/accepted_hits,bam
```

Figure 2.5: The script written to assemble a transcriptome for each sample using Cufflinks.

```
Cuffmerge_blood.sh
 #!/bin/bash
 # OGE parameters
 #$ -q xe-el6
#$ -N CuffMerge_Blood
#$ -e /no_backup/xe/ahobbs/DE/e,cm,logs
#$ -o /no_backup/xe/ahobbs/DE/o,cm,logs
#$ -V
 #$ -m abe
#$ -M angela,hobbs@crg,eu
 #$ -t 1
 #$ -pe smp 8
#$ -l h_rt=30:00:00
#$ -l virtual_free=40G

source /users/xe/ahobbs/,bash_profile
source /users/xe/ahobbs/,bashrc

/users/GD/tools/cufflinks/cufflinks-2,2,1,Linux_x86_64/cuffmerge -o /no_backup/xe/ahobbs/DE --num-threads 8 -g
/db/igenomes/Homo_sapiens/Ensembl/GRCh37/Annotation/Genes/genes,gtf -s
/db/igenomes/Homo_sapiens/Ensembl/GRCh37/Sequence/Bowtie2Index/genome,fa /no_backup/xe/ahobbs/samples/assemblies_blood,txt
```

Figure 2.6: **The Cuffmerge script written to merge all the resulting transcriptome fragments of each sample together with the reference transcriptome annotation.**

The merged file was then quantified by cuffdiff which is a separate program that is included in the cufflinks package. Cuffdiff calculated differential gene expression i.e. the expression between our case and control groups and also tested the statistical significance of each observed change in the expression between them. The results were given in a set of tabular files. Differential expression was considered significant depending on whether the p-value is greater than the FDR after Benjamini-Hochberg correction for multiple-testing (Mutryn et al., 2015). The output file generated by cuffdiff was saved in an excel format for analysis.

```
Cuffdiff.placenta.final.sh
 #!/bin/bash
 # OGE parameters
 #$ -q xe-el6
#$ -N cuffdiff_placenta
#$ -e /no_backup/xe/ahobbs/cuffdiffplacentafinal/e,CDplacentafinal,logs
#$ -o /no_backup/xe/ahobbs/cuffdiffplacentafinal/o,CDplacentafinal,logs
#$ -V
 #$ -m abe
#$ -M angela,hobbs@crg,eu
 #$ -t 1
 #$ -pe smp 8
#$ -l h_rt=72:00:00
```

```
#$ -l virtual_free=60G



aligndir=/no_backup/xe/ahobbs/alignment

/users/GD/tools/cufflinks/cufflinks-2,2,1,Linux_x86_64/cuffdiff -o /no_backup/xe/ahobbs/cuffdiffplacentafinal -p 8 -L Controls,Cases --library-type fr-
firststrand /db/igenomes/Homo_sapiens/Ensembl/GRCh37/Annotation/Genes/genes,gtf
$aligndir/0002/accepted_hits,bam,$aligndir/1090/accepted_hits,bam,$aligndir/0017/accepted_hits,bam,$aligndir/0018/accepted_hits,bam,$aligndir/
0002/accepted_hits,bam,$aligndir/1090/accepted_hits,bam
$aligndir/0006/accepted_hits,bam,$aligndir/0007/accepted_hits,bam,$aligndir/0013/accepted_hits,bam,$aligndir/0006/accepted_hits,bam,$aligndir/
0007/accepted_hits,bam,$aligndir/0013/accepted_hits,bam
```

Figure 2.7: The Cuffdiff script. This script was written to extract differential gene expression sequences from the blood and placenta dataset.

The number of RNA-seq reads generated from a transcript is directly proportional to the relative abundance of that transcript in the sample and because cDNA fragments are generally size-selected as part of library construction, longer transcripts produce more sequencing fragments than shorter transcripts. In order to determine the correct expression level of each transcript, cufflinks must count the reads that map to each transcript and then normalize this count by each transcript's length. The commonly used fragments per kilobase of transcript per million mapped fragments (or FPKM, also known as RPKM in single ended sequencing experiments) is used to normalization expression levels for different genes and transcripts (Trapnell et al., 2012). Figure 2.7 shows the script written to run the cuffdiff command to extract differential expression gene sequences.

CummeRbund (http://compbio.mit.edu/cummeRbund) is a powerful plotting tool which was used to create commonly used expression plots such as volcano, scatter and box plots, cummeRbund transforms cufflinks output files into R objects suitable for analysis with a wide variety of other packages available within the R environment. The cuffdiff output file was used as an input for cummeRbund.

2.6 Candidate gene selection

Pairwise comparisons between GDM cases and controls were carried out to identify genes that displayed significant ($p<0.05$) differential expression (log2 fold change> or <2) in the blood and placenta dataset. Corrections for multiple testing were performed using the

Benjamini-Hochberg correction method. Cuffdiff produced a list of thousands of genes which display differential expression between the case and control groups in both tissues. A filtering process was required to identify genes whose differential expression was statsitically significant (qvalue<0.05). This helped to reduce this list to a smaller workable number of potential candidate genes which may play a role in the development of gestational diabetes.

2.6.1 Functional annotation of genes

For an interpretation of the biological functions and molecular processes of the genes that show a statistically significant level of differential expression between patient and control groups, Gene Ontology (GO) analysis using the PANTHER-v8.1 (Protein Analysis Through Evolutionary Relationships) Classification System (http://pantherdb.org/) and DAVID (Database for Annotation, Visualisation and Integrated Discovery) (https://david.ncifcrf.gov/) were used. PANTHER uses the binomial statistics tool to compare the input gene list to a reference list (NCBI: Homo sapiens genes) to determine the statistically significant over-representation of functional groups of genes. Pathway level analysis of gene expression data was performed by Gene Set Enrichment Analysis (GSEA). The large gene lists were manually analysed to identify genes which displayed consistent significant differential expression in each of the individual GDM cases but not in any of the controls. The filtering process enabled the identification of genes which are likely to be strong functional and biological candidates. A comprehensive literature search was done using OMIM, Google scholar and PubMed on each of the selected genes to identify a possible role in the development of GDM. Search terms used included, diabetes, diabetes susceptibility gene, gestational diabetes, insulin resistance, glycaemic control, biomarker and glucose metabolism. Using all of the above information, potential candidate genes were selected for validation.

2.7 Validation of RNA sequencing data using TaqMan probe assays

2.7.1 cDNA Synthesis

Reverse transcription was performed on an input of 250 ng of total RNA per sample using the High Capacity RNA-to-cDNA kit (Thermo-Fisher Scientific). The reactions for each RNA sample as well as control reactions (do not contain reverse transcriptase (NRT)) were set up according to the manufacturer's instructions. The cDNA was quantified using the Nanodrop, normalised

and stored at -20°C until needed (The protocol for reverse transcription is available in Appendix F).

2.7.2 Preparing the TaqMan Gene Expression Assays

Equal concentrations of each of the samples (100 ng of total cDNA) was used to accurately measure gene expression using the TaqMan assay method. Each assay (specific for each target gene) was pre-designed and readily available from Thermo-Fisher Scientific (The Assay number for each gene assay is listed in Appendix F). The expression of each of the target genes was normalised to the expression of three housekeeping genes namely, *ACTB*, *HPRT1* and *RPLP0*. For each assay, every sample, including a calibrator, was run in triplicate, a no-reverse transcriptase and a no-amplification control was included. The TaqMan gene expression assays were run in a 384-well plate format in the ABI-7900HT Real-Time PCR machine. (The protocol for the TaqMan qPCR is available in Appendix F).

2.7.3 qRT-PCR data analysis

The data generated were analysed using the RQ (relative quantification) software on the ABI 7900HT Real-Time PCR machine. Prior to analysis, specific parameters for the analysis of the data were configured following the instruction manual for this purpose. This included the selection of endogenous control samples, adjusting the baseline and threshold values as well as viewing the amplification and expression plots. The data were then exported for further analysis using the R/Bioconductor packages ReadqPCR and NormqPCR (Perkins et al., 2012).

2.8 Promoter region methylation analysis of candidate genes

The methylation status of selected CpG islands in the promoter regions of the *G6PD (*CpG Island 115375**),** *TKT (*CpG Island 110332), *IGFBP-1 (*CpG Island 113146); *IGFBP-2* (CpG Island 108855) and *IGFBP-6* (CpG Island 103158) genes were examined by methylation-specific PCR. This was carried out in the whole blood samples and placental tissue using the EpiTect Methyl II PCR assay (Qiagen) procedure. This method is based on the quantitative detection of remaining input DNA within a sample population after treatment with a methylation-sensitive (MSRE) and a methylation-dependent (MDRE) restriction enzyme. Primers were designed by an optimized computer

algorithm to ensure that the amplicon contains cutting sites for both methyl-sensitive and methyl-dependent enzymes and are specifically designed for analyzing the DNA methylation status of CpG islands using restriction enzyme digestion (SABiosciences-Qiagen).

Briefly, 2 µg input genomic DNA was aliquoted into four equal portions and subjected to mock (no enzyme), methylation-sensitive (MSRE), methylation-dependent (MDRE), and double (MSRE and MDRE) restriction endonuclease digestion. The product of the mock (no enzyme) digestion represents the total amount of input DNA for real-time PCR detection. In the methylation-sensitive digestion (Ms) reaction, the MSRE will digest unmethylated and partially methylated DNA. Consequently, the amount of DNA remaining in this reaction represents the fraction of fully methylated DNA within the sample population. The remaining hypermethylated DNA, DNA in which all CpG sites are methylated, will be detected by real-time PCR. In contrast, the fraction of unmethylated DNA is determined by the methylation-dependent digestion (Md) reaction as the MDRE will digest methylated DNA. The remaining unmethylated DNA was detected by real-time PCR. The amount of remaining input DNA in each digest is then normalized to the total amount of input DNA, which is determined by the mock-treated DNA fraction. Both enzymes are present in the double digestion (Msd) reaction so methylated and unmethylated DNA will be digested. This reaction measures the background and the fraction of input DNA vulnerable to enzyme digestion. To ensure restriction enzyme efficiency, two controls were added for each assay, Sensitive Enzyme Control (SEC) and Dependent-Enzyme Control (DEC).

After digestion, the enzyme reactions were mixed directly with qPCR master mix and pre-designed gene-specific primer mixes. Real-time PCR was carried out using specified cycling conditions (Full protocol can be found in Appendix G). Finally, the raw ΔC_T values were pasted into the data analysis spreadsheet (provided by Qiagen) which automatically calculates the relative amount of methylated and unmethylated DNA fractions. Unmethylated represents the fraction of input genomic DNA containing no methylated CpG sites in the amplified region of a gene. Methylated represents fraction of input genomic DNA containing two or more methylated CpG sites in the targeted region of a gene. Comparisons of gene promoter levels between 2 groups (cases blood vs controls blood, cases placenta vs controls placenta; cases blood vs cases placenta; controls blood vs controls placenta) were determined using two tailed

t-tests. A paired test was used when comparing the blood values with the placental values ie data derived from the same individual. When comparing the cases with the controls a non-paired Student's t test was used. The p-values of less than 0.05 were considered statistically significant.

2.9 Inferring correlation networks using Spearman's Rank correlation (SRC)

In this study it is important to determine whether there is a correlation between the mRNA expression levels of a gene in either the blood or placental samples and other measured variables such as maternal glucose levels, maternal BMI and foetal birth weight. These correlations were determined using the SRC. Spearman's correlation is a non-parametric statistical evaluation that is used to study the strength of a relationship between two variables (X and Y). The Spearman's coefficient is denoted by Rs (or ρ) and the value of Rs ranges from -1 to 1. A perfect Spearman correlation (Rs = 1 or Rs = -1) indicates a monotonic relationship between the two ranked variables. The sign marks the direction of the correlation: $\rho > 0$ (positive correlation) if Y tends to increase when X increases and $\rho < 0$ (negative correlation) if Y tends to decrease when X increased. A p-value for the association is also reported (a $p \leq 0.05$ is considered to indicate statistical significance). Spearman's Rank correlation analysis was performed using Intellectus Statistics (http://www.intellectusstatistics.com) and the results were validated in Excel.

CHAPTER THREE

3. RESULTS

3.1 The identification and selection of study participants

All GDM cases were diagnosed based on a fasting blood glucose level of ≥ 5.1 mmol/L; a 1-hour blood glucose level of ≥ 10.0 mmol/L and/or a 2-hour blood glucose level of ≥ 8.5 mmol/L. If only one or more values were equal to or exceeded the diagnostic thresholds recommended by the IADPSG criteria GDM was diagnosed. Women who had normal blood glucose levels at each measured time frame (fasting ≤ 5.0 mmol/L; 1-hour ≤ 8.0 mmol/L and/or 2-hour ≤ 8.0 mmol/L), were selected as controls. Table 3.1 lists the clinical and biochemical characteristics of the maternal group. The OGTT was performed at between 29 to 33 weeks gestation. Each measurement indicative of GDM is bolded in red. The difference in blood glucose levels at fasting and 1-hr is significantly different (p=0.003 and p=1.93 x 10^{-5}, respectively) between the cases and controls. The blood glucose levels at 2-hr post OGTT was not found to be significantly different between the two biological groups (p=0.08). However, it was believed that the samples making up the two biological groups were diverse enough for the objectives of this study (Figure 3.1). Figure 3.1 illustrates that there is a clear separation between the blood glucose levels at each measured time point between the case and control samples used in this study. The OGTT values in the GDM group (cases) were higher at each measured point when compared to the values of the control group.

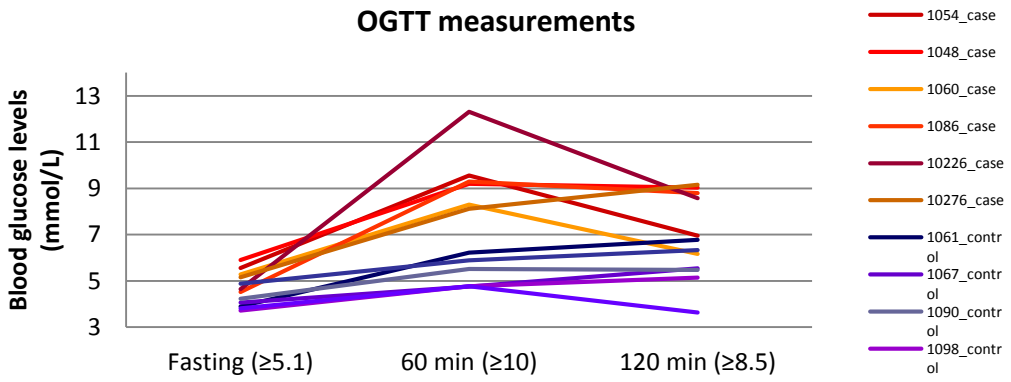


Figure 3.1: **The Oral Glucose Tolerance Test (OGTT) results for each case and control sample.** The red and orange lines indicate GDM cases and the blue and purple lines represent controls.

The Body Mass Index (BMI) of the cases did not differ significantly (p=0.06) to that of the controls. The case group had a slightly higher overall BMI (37.9 kg/m^2) in comparison to the control group (30.8 kg/m^2). Women with gestational diabetes tend to have an abnormally high weight gain during pregnancy. To determine whether this was seen in the study participants, each women's weight was recorded at each prenatal visit (at <14 weeks, 14-18 weeks, 19-23 weeks, 24-28 weeks, 29-33 weeks and 34-38 weeks). The overall average weight gain in the case group was 11.4 kg which was slightly higher than the average weight gain in the control group which was 8.1 kg. This weight gain was not significantly different (p=0.052) between the GDM and control groups (Figure 3.2 and Table 3.1). It was not possible to match the cases and controls for age and BMI, but we attempted to align the two groups as closely as possible. The women with GDM had an average age of 31.3 years compared to 26.7 years for the NGT group. Half the GDM group were over 35 years of age, whereas only one control was. This is not surprising given that age is a risk factor for GDM, but for reasons not yet known. All women in the study were HIV negative, had no history of other diseases, were non-smokers and were not receiving any medication (with specific reference to insulin) at the time blood samples were taken. Therefore, the relevant differentiating variable between the biological groups is the presence or absence of gestational diabetes.

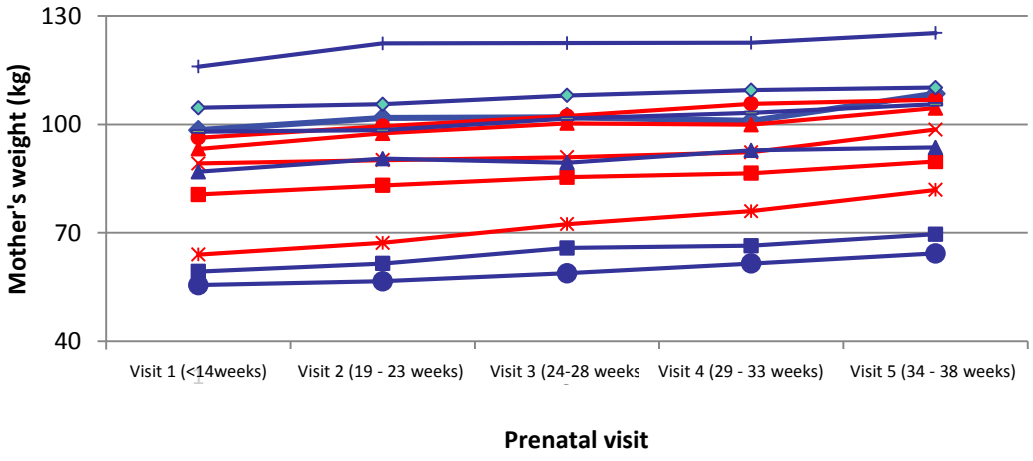


Figure 3.2: A graphical illustration of the women's weight gain during pregnancy. The red lines indicate case subjects and blue lines indicate control subjects.

Table 3.1: **The clinical and biochemical characteristics of the maternal study group.** The values bolded in red are those equal to or exceeding the diagnostic threshold that is indicative of GDM.

Lab code	Glucose levels			BMI (kg/m2) Visit 1 (<14 weeks)	Weight (kg) Visit 2 (14 - 18 weeks)	Weight (kg) Visit 3 (19 - 23 weeks)	Weight (kg) Visit 4 (24 - 28 weeks)	Weight (kg) Visit 5 (29 - 33 weeks)	Weight (kg) Visit 6 (34 - 38 weeks)	Weight gain (kg)	Age (years)	HIV Status	Previous pregnancies
	Fasting (>5.1mmol/l)	1 hr OGTT* (>10mmol/l)	2 hr OGTT* (>8.5mmol/l)										
GDM													
Case0_1054	**5.56**	9.65	6.95	35.4	98.4	101.8	102.1	101.1	108.5	10.1	39	Neg	1
Case1_1048	**5.9**	9.19	**9.03**	36.5	80.6	83.1	85.4	86.5	89.7	9.1	23	Neg	0
Case2_1060	**5.28**	9.38	4.17	36.1	93.2	97.5	100.3	99.9	104.5	11.3	25	Neg	0
Case3_1086	4.52	9.64	**8.91**	36.2	89.2	90.1	90.9	92.3	98.6	9.4	35	Neg	1
Case4_10225	4.64	**10.75**	6.71	48.5	116.2	122.4	122.5	122.6	125.3	17.9	31	Neg	1
Case5_10276	**5.16**	8.14	**9.15**	35.0	96.3	99.5	102.3	105.7	106.8	10.5	35	Neg	1
Group average	**5.2**	**9.5**	**7.5**	**37.9**	**87.0**	**89.9**	**92.2**	**93.6**	**98.3**	**11.4**	**31.3**	**N/A**	**0.7**
NGT													
Control0_1061	3.9	7.22	6.77	24.0	64.0	67.2	72.4	76.2	81.9	9.3	25	Neg	0
Control1_1067	4.06	4.76	5.54	23.4	55.5	56.6	58.8	61.5	64.3	8.8	22	Neg	2
Control2_1090	4.23	5.52	5.48	35.3	98.1	98.4	101.6	103.2	105.6	7.6	27	Neg	0
Control3_1094	3.71	4.77	5.14	29.7	89.3	91.5	95.8	96.4	99.6	10.3	34	Neg	1
Control4_1107	4.88	5.89	7.33	38.6	104.6	105.6	108.2	109.5	110.2	5.6	22	Neg	1
control5_1087	3.82	4.77	3.64	34.0	86.9	90.5	89.4	92.8	93.6	6.7	30	Neg	0
Group average	**4.1**	**5.5**	**5.7**	**30.8**	**86.7**	**89.2**	**91.0**	**92.7**	**94.8**	**8.1**	**26.7**	**N/A**	**0.7**
Significance	**P = 0.003**	**P=0.00009**	**P=0.08**	**P=0.06**						**P=0.052**			

Abbreviations: BMI: Body Mass Index; GDM: Gestational Diabetes Mellitus; NGT: Normal Glucose Tolerance; OGTT: Oral Glucose Tolerance Test; Neg: Negative

*OGTT done at approximately 24 to 28 weeks gestation

With regards to the newborns (Table 3.2), there was no significant difference between gestational age at birth (38.5 vs 38.7 weeks, p=0.41) and birth weight (3.40 vs 3.38 kg; p=0.85) between exposed and unexposed placenta respectively.

Table 3.2: **Clinical characteristics of the foetal (newborn) study group**

Lab code	Sex	Birth Weight (kg)	Length (cm)	Gestational Age (weeks)
GDM (mother)				
Case0_1054	FEMALE	3.68	44.0	39
Case1_1048	FEMALE	3.12	46.2	39
Case2_1060	FEMALE	3.43	44.6	38
Case3_1086	FEMALE	3.28	48.1	37
Case4_10225	FEMALE	3.22	42.1	38
Case5_10276	FEMALE	3.68	45.2	40
Group average	**N/A**	**3.40**	**45.0**	**38.5**
NGT (mother)				
Control0_1061	FEMALE	3.86	54.0	37
Control1_1067	FEMALE	3.16	53.0	39
Control2_1090	FEMALE	3.50	48.8	39
Control3_1094	FEMALE	2.90	50.4	38
Control4_1107	FEMALE	3.87	46.3	39
Control5_1087	FEMALE	2.89	44.5	40
Group average	**N/A**	**3.36**	**49.5**	**38.7**
Significance		**P=0.85**	**P=0.41**	

Abbreviations: GDM: Gestational Diabetes Mellitus; NGT: Normal Glucose Tolerance

3.2 Nucleic acid extraction from blood and placental tissue

3.2.1 Deoxyribonucleic acid (DNA)

The Nanodrop 1000 Spectrophotometer was used to measure the quantity (ng/µl) and quality (260/280 and 260/230 measurements) of the DNA extracted from whole blood and placental tissue. Table 3.3 and Table 3.4 lists these measurements for blood and placenta tissue respectively. Across both blood (b) and placental (p) samples, the DNA quality and quantity did not differ significantly. The concentration ranged from 302 ng/µl – 855.9 ng/µl for the blood samples and from 31 ng/µl – 162.3 ng/µl for the placental samples. The difference in the concentration range between the blood and placenta samples is noticeable and may be due to the different extraction kits/methods that were used for the different tissues (salting out versus a Qiagen kit). For the blood samples, the ratio of absorbance at 260/280 ranged from 1.79– 1.88 and the 260/230 values ranged from 2.25 – 2.53. For the placenta samples, the 260/280 values ranged from 1.82 – 2.03 and the 260/230 values ranged from 2.05 – 2.63. Overall, the DNA extracted from whole blood and placental tissue was of good quality and efficient quantity for further downstream analysis.

Table 3.3: **The quality and quantity of the DNA extracted from the blood samples**

DNA from Blood (b) Lab Code		Concentration (ng/µl)	Quantity (µg)*	UV 260/280	UV 260/230
1b	**10276**	765.9	153.2	1.87	2.25
2b	**1107**	164.0	32.8	1.79	2.38
3b	**1086**	660.5	132.1	1.87	2.50
4b	**1067**	617.1	123.4	1.90	2.29
5b	**1060**	403.8	80.76	1.79	2.45
6b	**1061**	265.4	53.08	1.80	2.53
7b	**10225**	167.3	33.46	1.78	2.37
8b	**1094**	690.4	138.1	1.87	2.45
9b	**1090**	904.5	180.9	1.85	2.33
10b	**1048**	242.7	48.54	1.80	2.53
11b	**1054**	184.9	36.98	1.88	2.47
12b	**1087**	215.7	43.14	1.79	2.60

* Total volume eluted was 200 µl. Quantity (µg) calculated as concentration (ng/µl) * 200 µl/1000

Table 3.4: **The quality and quantity of the DNA extracted from placental tissue**

DNA from Placenta (p) Lab Code		Concentration (ng/µl)	Quantity (µg)*	UV 260/280	UV 260/230
1b	**10276**	86.8	17.4	1.95	2.44
2b	**1107**	119.5	23.9	2.02	2.63
3b	**1086**	83.0	16.6	1.96	2.46
4b	**1067**	110.7	22.1	1.98	2.14
5b	**1060**	141.6	28.3	1.85	2.21
6b	**1061**	81.4	16.3	1.82	2.05
7b	**10225**	59.6	11.9	1.99	2.11
8b	**1094**	159.9	31.9	1.90	2.45
9b	**1090**	162.2	32.4	1.92	2.37
10b	**1048**	113.4	22.7	1.82	2.11
11b	**1054**	184.9	36.9	1.88	2.47
12b	**1087**	108.5	21.7	2.03	2.46

* Total volume eluted was 200 µl. Quantity (µg) calculated as concentration (ng/µl) * 200 µl/1000

3.2.2 Ribonucleic Acid (RNA)

The Agilent 2100 BioAnalyzer was used to measure the quality (RNA integrity number (RIN) and rRNA ratio measurements) and quantity (concentration ng/µl) of the RNA extracted from whole blood and placental tissue. The Nanodrop 1000 Spectrophotometer was used to obtain the 260/280 and 260/230 measurements which are also indicative of RNA purity. Tables 3.5 and 3.6 list the quality and quantity measurements of the RNA extracted from whole blood and placental samples respectively. The concentration of total RNA extracted from whole blood tissue ranged from 90.9 ng/µl – 455.8 ng/µl and from 84.2 ng/µl – 665.4 ng/µl for the RNA extracted from the placental tissue. The differences in the concentration range between the different tissues is relatively small and may be due to the different extraction methods/kits used. The RIN and rRNA ratios indicate sample quality, RIN values >7 and rRNA ratios >1.7 indicate good quality RNA. The RIN values and rRNA ratio for the blood samples ranged from 7.3 – 8.5 and 1.7 – 2.2 respectively. The RIN values and rRNA ratio for the placenta samples ranged from 7 – 8.3 and 1.7 – 2.4 respectively (Table 3.5 and Table 3.6). The 260/280 ratio of absorbance values for the blood samples ranged from 2.02 – 2.08 and the 260/230 values ranged from 2.05 – 2.47. For the placenta samples the

260/280 values ranged from 2.02 – 2.11 with the 260/230 values ranging from 2.05 – 2.57. Overall, the extracted RNA was of good quality and efficient quantity for further downstream analysis.

Table 3.5: The quality and quantity of the total RNA extracted from blood

RNA from Blood (B) Lab Code		Concentration (ng/µl)	Quantity (µg)*	UV 260/280	UV 260/230	RIN	rRNA ratio (28s/18s)
1b	**10276**	153.9	13.08	2.08	2.11	8.1	1.7
2b	**1107**	90.9	7.73	2.02	2.05	8.0	1.7
3b	**1086**	200.5	17.04	2.05	2.05	7.7	1.7
4b	**1067**	278.3	23.66	2.05	2.33	8.0	1.7
5b	**1060**	455.8	38.74	2.04	2.48	8.2	1.7
6b	**1061**	300.7	25.56	2.05	2.46	8.2	1.8
7b	**10225**	379.1	32.22	2.04	2.13	7.6	1.8
8b	**1094**	262.4	22.3	2.05	2.47	8.2	1.8
9b	**1090**	280.3	23.83	2.03	2.47	8.5	1.7
10b	**1048**	300.3	25.53	2.05	2.29	7.4	1.9
11b	**1054**	220.3	18.73	2.02	2.35	7.3	2.0
12b	**1087**	101.8	8.65	2.07	2.38	7.3	2.2

* Total volume eluted was 85 µl. Quantity (µg) calculated as concentration (ng/µl) * 85 µl/1000

Table 3.6: The quality and quantity of the total RNA extracted from placental tissue

RNA from Placenta (p) Lab Code		Concentration (ng/µl)	Quantity (µg)**	UV 260/280	UV 260/230	RIN	rRNA ratio (28s/18s)
1p	**10276**	86.3	4.75	2.11	2.11	7.1	1.7
2p	**1107**	84.2	4.63	1.98	2.05	7.0	1.7
3p	**1086**	330.2	55.17	2.11	2.32	8.0	1.7
4p	**1067**	665.4	29.06	2.05	1.94	7.6	1.8
5p	**1060**	545.1	46.43	2.1	2.25	7.7	1.7
6p	**1061**	527.4	31.48	2.03	2.33	7.5	1.9
7p	**10225**	379.1	16.73	2.04	2.13	7.6	1.9
8p	**1094**	262.1	10.64	2.05	2.57	8.2	2.4
9p	**1090**	493.7	37.77	2.14	2.16	7.6	2.0
10p	**1048**	400.0	36.05	2.14	2.18	7.0	2.2
11p	**1054**	509.0	43.5	2.02	2.35	7.3	1.8
12p	**1087**	471.7	31.61	2.08	2.25	8.3	1.7

** Total volume eluted was 55 µl. Quantity (µg) calculated as concentration (ng/µl) * 55 µl/1000

3.2.3 Globin mRNA removal from whole blood RNA

Due to the large amounts of globin mRNA in whole blood samples which could potentially affect the outcome of sequencing results, the RNA extracted from whole blood was subjected to a globin removal step (this process is not necessary for the RNA samples extracted from placental tissue). This removal step may slightly decrease the quality and quantity of the RNA but may increase the identification of low expressing transcripts. It is necessary to re-measure the quality of the RNA once the mRNA globin has been removed.

The quality and quantity measurements of the RNA after globin mRNA removal are listed in Table 3.7. Only a portion (3 µg) of the total extracted RNA samples was used as the input for this globin removal step. The 'output' indicates what percentage of RNA was lost in the removal step and this varied substantially across the samples (0.87 – 2.92 µg). An output/input ratio of 1 indicates that no RNA was lost during the removal process where a ratio of 0 indicates that the entire RNA sample was lost. The output/input ratio ranged from 0.29 – 0.97. Although some samples lost >50 % of input RNA (sample 1b and 2b), the quantity and quality of all these samples was sufficient to continue with mRNA library preparation.

According to the literature, the RIN and rRNA ratio of RNA should decrease after a globin removal step (Krjutskov et al., 2016, Liu et al., 2006). Figure 3.3 illustrates an overall decrease (in approximately 1 unit for each sample) in the rRNA ratio for the RNA samples but an unexpected increase (in approximately 1 unit for each sample) in the RIN. The RIN measurement was repeated on two different BioAnalyzer machines but had the same result each time. The rRNA ratios after globin removal ranged from 1-1.8 (still acceptable for sequencing according to Illumina) and the RIN ranged from 7.2 – 9.1.

Table 3.7: The quality and quantity of the RNA AFTER the globin mRNA removal step

Blood RNA Samples	Concentration (ng/µl)	Input Quantity (µg) Total RNA	Output Quantity (µg)	Ratio (Output µg/ Input µg)	UV 260/280	UV 260/230	RIN	rRNA ratio (28S/18S)
1b	34.9	3	0.87	0.29	1.87	0.30	9.1	1.6
2b	46.3	3	1.16	0.39	1.96	0.49	9.0	1.7
3b	77.9	3	1.95	0.65	2.04	0.85	8.7	1.2
4b	93.6	3	2.34	0.78	2.03	0.81	8.9	1.5
5b	103.9	3	2.6	0.87	2.01	0.63	8.7	1.6
6b	86.8	3	2.17	0.72	1.99	0.89	8.4	1.0
7b	91.0	3	2.28	0.76	2.02	0.79	8.6	1.6
8b	64.0	3	1.6	0.53	2.01	0.61	8.8	1.7
9b	100.4	3	2.51	0.84	1.97	0.80	9.0	1.7
10b	116.7	3	2.92	0.97	1.99	1.05	7.2	1.0
11b	90.4	3	2.25	0.75	2.02	0.95	8.4	1.8
12b	85.3	3	2.13	0.71	2.01	0.72	8.0	2.0

A) B)

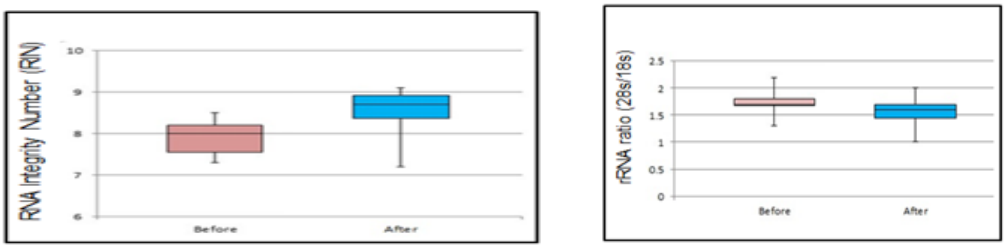


Figure 3.3: The average change in RNA integrity by A) RIN measurement and B) rRNA ratio after the globin mRNA step.

3.3 cDNA library preparation using the Illumina TruSeq Stranded mRNA preparation kit

The quality and quantity of the RNA extracted from whole blood and placental tissue met the Illumina quality control guidelines (Conesa et al., 2016, Trapnell et al., 2013, Trapnell et al., 2012). All samples were therefore used for cDNA/mRNA library preparation for RNA sequencing.

3.3.1 Library Validation

The concentrations of the libraries varied significantly (1.51 – 28.23 ng/µl), however, even the library with the smallest concentration (1.51 ng/µl) was sufficient for sequencing. According to the Illumina protocol, the size of the libraries should be >260 bp; these ranged from 296 – 361 bp. No samples failed library preparation and each library was used for sequencing.

Table 3.8: The quantity and size of the resulting mRNA libraries

mRNA library results BLOOD RNA		Concentration (ng/µl)	Size (bp)	Molarity (nmol/l)	mRNA library results PLACENTAL RNA		Concentration (ng/µl)	Size (bp)	Molarity (nmol/l)
1b	**10276**	1.5	337	7.2	**1p**	**10276**	24.1	336	115.0
2b	**1107**	30.0	337	146.6	**2p**	**1107**	20.4	320	101.4
3b	**1086**	7.5	296	38.5	**3p**	**1086**	19.0	322	94.2
4b	**1067**	21.4	296	109.8	**4p**	**1067**	28.2	333	136.3
5b	**1060**	7.1	312	35.9	**5p**	**1060**	6.1	337	32.7
6b	**1061**	11.4	337	54.5	**6p**	**1061**	8.3	316	39.9
7b	**10225**	26.7	324	132.4	**7p**	**10225**	20.7	335	99.5
8b	**1094**	12.3	337	58.7	**8p**	**1094**	19.7	327	96.1
9b	**1090**	16.0	361	72.2	**9p**	**1090**	9.3	337	50.0
10b	**1048**	5.0	296	28.2	**10p**	**1048**	0.6	337	3.4
11b	**1054**	13.9	359	63.1	**11p**	**1054**	15.8	326	73.8
12b	**1087**	18.8	322	92.8	**12p**	**1087**	21.9	315	110.8

3.3.2 Whole genome-sequencing using the Illumina HiSeq 2000 Sequencing System

Sequencing in the HiSeq was set up to ensure excellent coverage of each sample: 4 samples per lane; 75 bp paired end (sequencing from the left and right). This criteria should produce more than 50,000,000 reads per sample. These millions of reads were mapped against a reference human genome (UCSC: GRCh38). Table 3.9 gives the sequencing statistics for each sample. The statistics include the following: **Input**: this is the total number of raw sequences generated through sequencing; **Mapped Left Reads**: this is the number of forward sequence reads mapped to the genome; **Mapped Right Reads**: this is the number of reverse sequence reads mapped to the genome and the **Alignment Score**: this is the percentage reflecting how well the total input raw reads aligned/mapped to the reference

genome, Illumina recommends an overall alignment score of >80 % and a Phred score of >30 (which indicates that base calling was done with 99.9 % accuracy and only 1:1000 bases were called with inaccuracy). The overall alignment/read mapping rate of the samples in this study ranged from 89.3 – 97.4 % and more than 96 % of the sequences were called with 99.99 % accuracy. For each sample, with the exception of 1048 placenta, more than 90 % of the produced reads mapped successfully to the genome. For sample 1048 placenta, only 80 % of the produced sequences mapped to the genome. This is still acceptable by Illumina guidelines.

Table 3.9: Sequencing statistics

	Input (No of reads)	Mapped Left Reads	% of input	Mapped Right Reads	% of input	Alignment (% overall read mapping rate)	Phred quality score (Q score>30)
1048placenta	52026636	41869010	90.5	40679970	88.2	89.3	96.0
1061placenta	58358710	53917568	92.4	52954040	90.7	91.6	96.1
1048blood	55158114	51531303	93.4	51174684	92.8	93.1	96.6
1054placenta	62219053	60113365	96.6	59576938	95.8	96.2	96.4
1060placenta	53042983	49708403	93.7	49277347	92.9	93.3	96.5
1090Placenta	72058148	70076414	97.2	69514836	96.5	96.9	96.7
1086blood	70811795	67064391	94.7	66704621	94.2	94.5	96.3
1067blood	69716379	67577952	96.9	66556776	95.5	96.2	96.0
10276blood	88050506	84099631	95.5	82784494	94.0	94.8	96.4
1094blood	85513343	83033919	97.1	82120902	96.0	96.6	96.1
1086placenta	93613852	91290170	97.5	89748267	95.9	96.7	95.9
1087placenta	85648415	83597848	97.6	82739970	96.6	97.1	97.0
1061blood	80754619	78394763	97.1	77565393	96.1	96.6	96.7
10225blood	86701746	84773781	97.8	83829540	96.7	97.2	96.8
1094placenta	106504113	103139807	96.8	102629475	96.4	96.6	96.4
1067placenta	99974533	96665446	96.7	95934572	96.0	96.3	95.6
1090blood	72043915	69224839	96.1	68632978	95.3	95.7	95.5
1107blood	82069631	79334105	96.7	78347937	95.5	96.1	96.4
10276placenta	88442275	85757815	97.0	85097904	96.2	96.6	96.2
1107placenta	61240854	59534955	97.2	58938622	96.2	96.7	96.4
1054blood	72312560	69857753	96.6	69455306	96.0	96.3	95.7
1060blood	79262465	76974845	97.1	76531105	96.6	96.8	96.0
1087blood	79620469	76964845	97.2	76963734	97.2	97.4	96.5
10225placenta	70396381	61189834	96.9	60458960	95.6	96.3	96.5

Figure 3.4 illustrates the number of sequences produced per sample and the number of these sequences that successfully mapped to the reference genome. Each of the samples produced over 50,000,000 reads, with the range being 52,026,636 to 106,504,113. The mapped left reads for each sample did not differ significantly to the mapped right reads indicating good sequencing coverage for each sample.

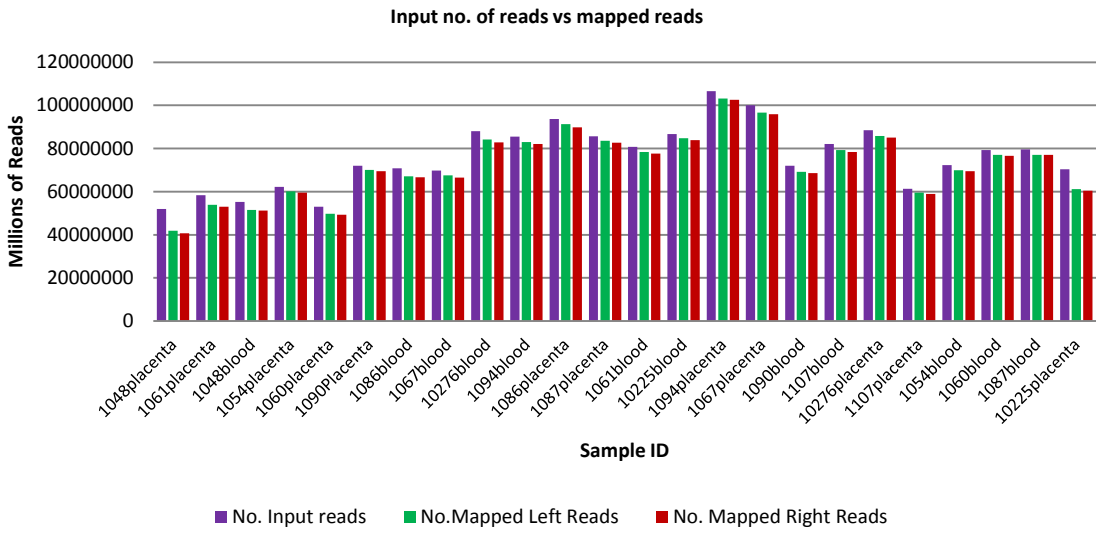


Figure 3.4: The total number of sequencing reads generated per sample. The number of reads which successfully mapped to the reference human genome (Genome Reference Consortium GRCh38 (hg38)) are also given.

3.3.2.1 FASTQC Quality Control of sequencing fragments

The quality of all of the sequences produced for each sample was evaluated using a program called FASTQC. FASTQC gives an indication of the following: **per base sequence** quality which is an overview of the range of quality values across all bases at each position; **per sequence quality score** which indicates if a subset of the sequences have universally low quality values; **per base sequence content** which plots out the proportion of each base position for which each of the four normal DNA bases has been called; **per base GC content** which plots out the GC content of each base position; **per sequence GC content** which measures the GC content across the whole length of each sequence and compares it to a

modelled normal distribution of GC content; **per base N content** which plots out the percentage of base calls at each position for which an N was called; **sequence length distribution** which generates a graph showing the distribution of fragment sizes in the file which was analysed and an indication of **duplicate sequences and overrepresented sequences**. A FASTQC file was generated for each sample. All sequences for each sample passed the QC evaluation and trimming of unreliable sequences was not necessary. All of the sequences produced for each sample were used for downstream differential expression analysis.

3.4. Analysing data generated from BLOOD (GDM vs NGT) comparisons

Once the aligned sequences passed the quality check, they were analysed using cufflinks and cuffdiff.

3.4.1 Cuffdiff output analysis

Table 3.10 lists the output generated by cuffdiff once it has been exported into an excel format. The test_id and gene_id are the Ensembl gene IDs and are unique identifiers describing the transcript, gene, primary transcript, or CDS being tested. The gene is identifed through its gene name. The genomic coordinate (locus) of the gene is also given. Sample_1 represents the control group and Sample_2 represents the case group. The test status is either given as **OK** (test successful), **NOTEST** (not enough alignments for testing), **LOWDATA** (too complex or shallowly sequenced), **HIDATA** (too many fragments in locus), or **FAIL**, when an ill-conditioned covariance matrix or other numerical exception prevents testing. Value_1 and value_2 represent the FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values of Sample_1 (controls) and Sample_2 (cases) respectively. The log2 fold change value is the (base 2) log of the fold change Sample_2(cases)/Sample_1 (controls). Therefore a positive log2 fold change represents an up-regulation of gene expression in the case group and a negative log2 fold change represents a down-regulation of the gene in the case group. The test_stat is the value of the test statistic used to compute

significance of the observed change in FPKM. The p-value is the uncorrected p-value of the test statistic and the q-value is the FDR-adjusted p-value of the test statistic. The significant column can be either “yes” or “no”, depending on whether the p-value is greater than the FDR after Benjamini-Hochberg correction for multiple-testing.

Table 3.10: **The output data obtained from running Cuffdiff on the blood dataset.** The same information is given for the placenta dataset.

test_id ENSG000000	gene_id ENSG000000	gene	locus	sample_1	sample_2	status
75213	75213	SEMA3A	7:83587658-84122040	Controls	Cases	OK
81985	81985	IL12RB2	1:67773046-67862583	Controls	Cases	OK
156113	156113	KCNMA1	10:78637354-79398353	Controls	Cases	OK
175445	156113	LPL	8:19759227-19824769	Controls	Cases	OK

value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
0.069	0.302	2.114	0.771	3.5E-04	2.1E-02	yes
0.808	2.438	1.592	1.490	4.5E-04	2.5E-02	yes
0.062	0.365	2.560	0.486	1.2E-03	4.8E-02	yes
0.547	1.408	1.362	1.473	5.0E-05	5.1E-03	yes

3.4.2 The filtering process used to reduce the number of potential candidate genes

A simple filtering process was used to identify genes which displayed significant differential expression (qvalue<0.05; log2 fold change of >2 or <-2) between cases (women who develop gestational diabetes) and controls (women with normal glucose tolerance during pregancy). The data were first sorted to only include genes with an “OK” status (N=17980). Only a few genes had the LOWDATA (N=32) and HIDATA (N=4) outcome and were subsequently excluded. The second filtering step removed all the genes that did not display significant differential expression (qvalue>0.05) between the cases and controls. This filtering step reduced the number of potential candidates. The level of statistical significance was increased to qvalue<0.05 and only genes with this level of significant differential expression were included. The total number of genes displaying highly significant (qvalue<0.05) differential expression beween the two groups was 440. The third filtering step focused on genes with highly significant (qvalue<0.05) differential expression

greater than 1 fold between the two biological groups. This gene list was further reduced by focussing on genes which had a differential expression greater than 2 fold between the two biological groups. These genes were targeted as potential candidates as they displayed the largest difference in expression between GDM cases and controls. This reduced the number of genes of interest to 60 (Figure 3.5).

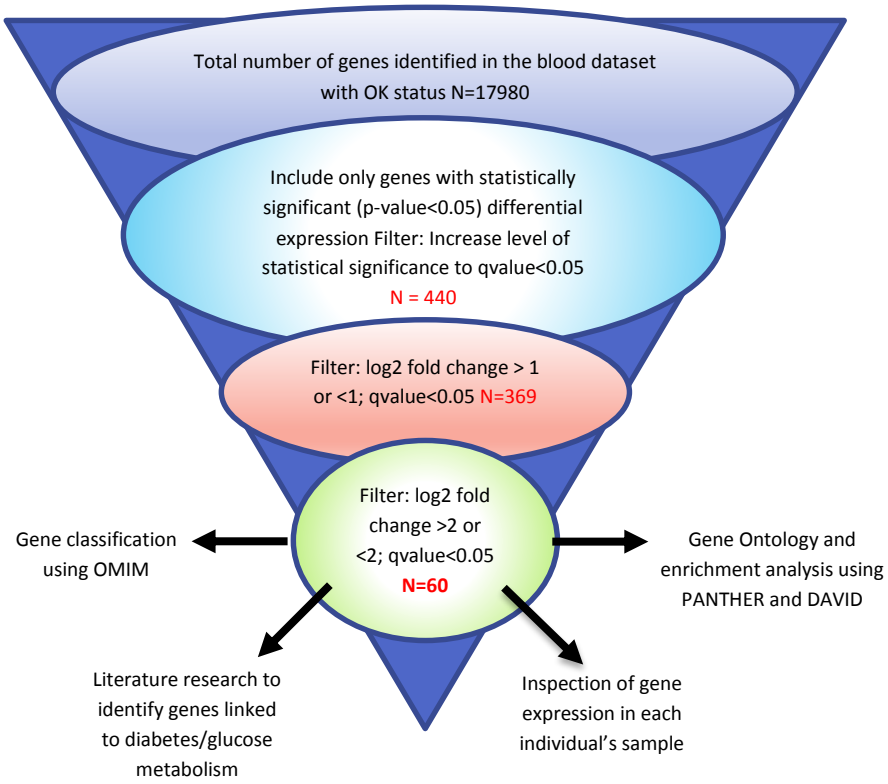


Figure 3.5: **A flow diagram depicting how the genes of interest from the blood dataset were selected for further analysis.**

3.4.2.1 The identification of genes displaying significant differential expression between case and control blood samples

The volcano plot (Figure 3.6) plots the relationship between the p-values of a statistical test and the magnitude of the difference in expression of 17980 genes. The $-\log_{10}$ p-values are plotted on the y-axis and the log2 fold change values are plotted on the x-axis. The red circles indicate the genes of interest that display both large-magnitude fold-change (x-axis)

as well as high statistical significance (-log10 of p-value, y-axis). The dashed red-line shows where qvalue = 0.05 with points above the line having qvalue < 0.05 and points below the line having qvalue > 0.05. This plot is colored such that the green points represent the genes with significant (qvalue>0.05) fold change of more than 1 (log2 < 1). The red dots represent genes with significant (qvalue>0.05) fold change of more than 2.

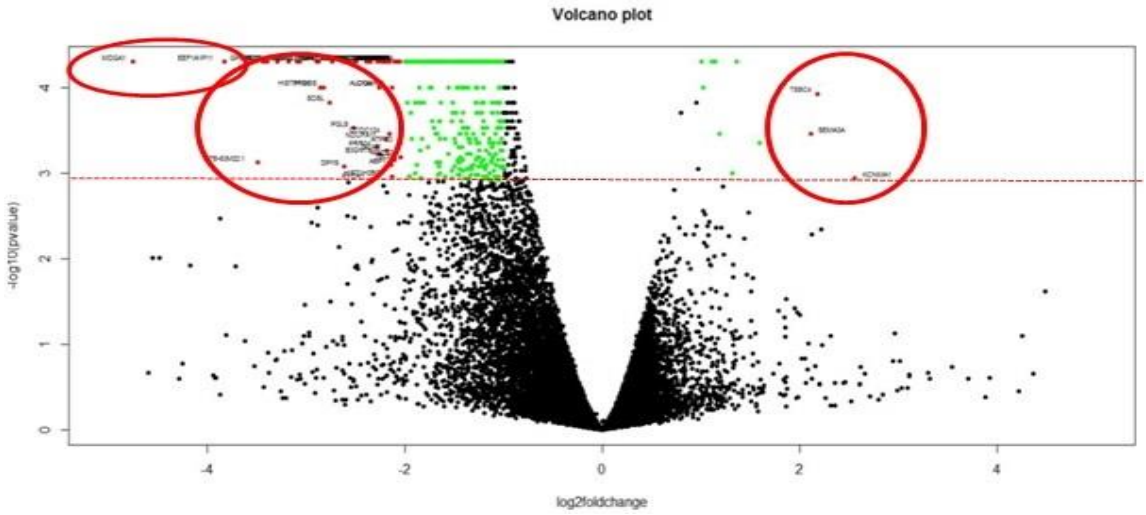


Figure 3.6: The volcano plot illustrating significant fold change differential expression for the blood data set comparing cases to controls. Volcano plot for the 17980 genes from the blood data set (log2 fold change >1; qvalue<0.05: green; log2 fold change >2; qvalue<0.05: red).

The scatter plot (Figure 3.7) illustrates the relationship between the corrected p-value (q-value) and the log2 fold change value of the genes which display the largest difference in expression and highest statistical significance (red dots in the volcano plot). The majority of the genes had a lower (-)(down regulated) gene expression in cases when compared to controls. Only 3 genes (*KCNMA1*; *SEMA3A* and *GSK3A*) had a higher (up regulated) expression in cases.

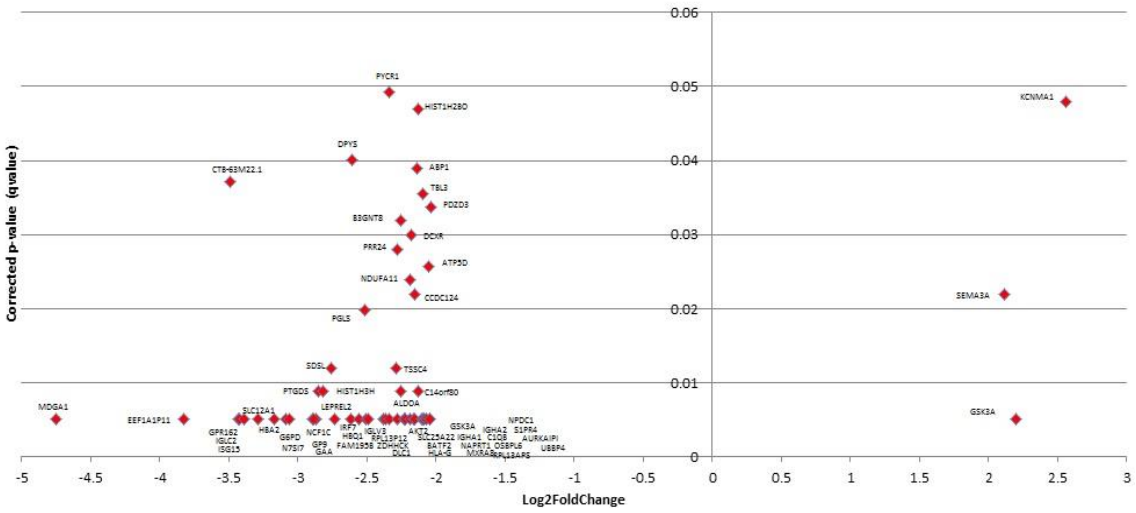


Figure 3.7: The Scatter Plot illustrating the genes which display significant fold change differential expression for blood dataset when comparing cases to controls. The scatter plot shows only the genes which have a log2 fold change >2 and <-2; qvalue<0.05 (corrected for multiple testing).

The 60 genes showing the most significant and largest differential expression in the case group (as highlighted in Figure 3.7) were used for further analysis. Table 3.11 lists these gene names, their location within the genome, the log2 fold change values and the corresponding p-value (showing level of significance) as well as the q-value. The q-values are the Benjamini-Hochberg corrected p-value measure (corrected for multiple testing).

Table 3.11: The top genes (N=60) displaying significant (qvalue<0.05) differential expression between cases and controls in the blood data set

Gene_id ENSG00000	Gene Symbol	Gene Name	Locus	log2(fold_change)	p_value	q_value
2726	**ABP1**	amiloride binding protein 1	7:150521714-150558592	-2.14	8.00E-04	0.0389
105221	**AKT2**	RAC-beta serine/threonine-protein kinase	19:40736223-40791443	-2.28	5.00E-05	0.0052
149925	**ALDOA**	Fructose-bisphosphate aldolase A	16:30064410-30081778	-2.26	1.00E-04	0.0089
99624	**ATP5D**	ATP synthase subunit delta. mitochondrial	19:1241748-1244824	-2.05	4.50E-04	0.0257
175756	**AURKAIP1**	Aurora kinase A-interacting protein	1:1309109-1310875	-2.05	5.00E-05	0.0052
177191	**B3GNT8**	BetaGal beta-1.3.N acetylglucosaminyltransferase8	19:41931263-41934635	-2.26	6.00E-04	0.0319
168062	**BATF2**	Basic leucine zipper transcriptional factor ATF-like 2	11:64755414-64764517	-2.23	5.00E-05	0.0052

185347	**C14orf80**	Uncharacterized protein C14orf80	14:105952653-105965912	-2.13	1.00E-04	0.0089
173369	**C1QB**	Complement C1q subcomponent subunit B	1:22979254-22988031	-2.10	5.00E-05	0.0052
7080	**CCDC124**	Coiled-coil domain-containing protein 124	19:18043824-18054800	-2.16	3.50E-05	0.0219
229119	**CTB-63M22.1**	Not Annotated	5:165809309-165809604	-3.49	7.55E-04	0.0372
169738	**DCXR**	Xylulose reductase	17:79993011-79995608	-2.18	5.50E-04	0.0299
164741	**DLC1**	Rho GTPase-activating protein 7	8:12940869-13373167	-2.34	5.00E-05	0.0052
147647	**DPYS**	Dihydropyrimidinase	8:105342551-105479281	-2.61	8.00E-04	0.0402
228502	**EEF1A1P11**	Not Annotated	1:96912485-96913874	-3.83	5.00E-05	0.0052
225663	**FAM195B**	Protein FAM195B	17:79780286-79791178	-2.51	5.00E-05	0.0052
160211	**G6PD**	Gluscose-6-phosphate dehydrogenase	X:153759605-153796782	-3.09	5.00E-05	0.0052
171298	**GAA**	Lysosomal alpha glucosidease	17:78075354-78093678	-2.88	5.00E-05	0.0052
169704	**GP9**	Platelet glycoprotein IX	3:128779609-128781249	-2.87	5.00E-05	0.0052
250510	**GPR162**	Not Annotated	12:6930710-6949018	-3.44	5.00E-05	0.0052
105723	**GSK3A**	Glycogen synthase kinase-3 alpha	19:42734337-42759309	3.20	5.00E-05	0.0052
188536	**HBA2**	Hemoglobin. alpha 2; hemoglobin. alpha 1	16:222845-223709	-3.17	5.00E-05	0.0052
86506	**HBQ1**	Hemoglobin subunit theta-1	16:230451-231180	-2.56	5.00E-05	0.0052
196331	**HIST1H2BO**	Histone H2B type 1-O	6:27861202-27861669	-2.13	1.00E-04	0.047
203813	**HIST1H3H**	Not Annotated	6:27777841-27778314	-2.86	1.00E-04	0.0089
146678	**IGFBP-1**	Insulin growth factor binding protein 1	7: 45888357 - 45893668	-2.22	5.00E-05	0.0052
211895	**IGHA1**	Ig alpha-1 chain C region	14:106173456-106175002	-2.19	5.00E-05	0.0052
211890	**IGHA2**	Ig alpha-2 chain C region	14:106053225-106054732	-2.10	5.00E-05	0.0052
211677	**IGLC2**	Ig lambda-2 chain C regions	22:23243155-23243617	-3.43	5.00E-05	0.0052
211662	**IGLV3-21**	Immunoglobulin lambda variable	22:23054173-23055688	-2.49	5.00E-05	0.0052
185507	**IRF7**	Interferon regulatory factor 7	11:612552-615999	-2.62	5.00E-05	0.0052
187608	**ISG15**	Ubiquitin-like protein ISG15	1:948802-949920	-3.39	5.00E-05	0.0052
156113	**KCNMA1**	Calcium-activated potassium channel subunit alpha-1	10:78637354-79398353	2.56	1.15E-03	0.048
53918	**KCNQ1**	Potassium voltage-gated channel subfam KQT member	11:2465913-2882798	-2.02	1.00E-04	0.0089
110811	**LEPREL2**	Prolyl 3-hydroxylase 3	12:6930710-6949018	-2.74	5.00E-05	0.0052
112139	**MDGA1**	MAM domain glycosylphosphatidylinositol anchor 1	6:37598454-37667082	-4.75	5.00E-05	0.0052
162576	**MXRA8**	Matrix-remodeling-associated protein 8	1:1288068-1297157	-2.16	5.00E-05	0.0052
147813	**NAPRT1**	Nicotinate phosphoribosyltransferases	8:144655659-144660819	-2.16	5.00E-05	0.0052
165178	**NCF1C**	Putative neutrophil cytosol factor 1C	7:74572444-74587848	-2.89	5.00E-05	0.0052
174886	**NDUFA11**	NADH dehydrogenase 1 alpha subcomplex. 11	19:5865836-5904017	-2.19	4.00E-03	0.0239
107281	**NPDC1**	Neural proliferation differentiation control protein 1	9:139933921-139940655	-2.07	5.00E-05	0.0052
79156	**OSBPL6**	Oxysterol binding protein-like 6	2:179059207-179264160	-2.09	5.00E-05	0.0052
172367	**PDZD3**	Na(+)/H(+) exchange regulatory cofactor NHE-RF4	11:119056165-119067479	-2.04	0.00065	0.0337
130313	**PGLS**	6-phosphogluconolactonase	19:17579577-17632097	-2.52	2.00E-03	0.0198

257704	**PRR24**	Not Annotated	19:47778141-47778979	-2.28	5.00E-03	0.028
107317	**PTGDS**	Prostaglandin-H2 D-isomerase	9:139871955-139880862	-2.82	1.00E-04	0.0089
183010	**PYCR1**	Pyrroline-5-carboxylate reductase 1. mitochondrial	17:79890259-79905477	-2.34	2.00E-03	0.0492
265150	**RN7SL2**	Not Annotated	14:50329270-50329567	-3.06	5.00E-05	0.0052
236552	**RPL13AP5**	Ribosomal protein L13a pseudogene 5	10:98510044-98510675	-2.09	5.00E-05	0.0052
215030	**RPL13P12**	Not Annotated	17:17286690-17287326	-2.39	5.00E-05	0.0052
125910	**S1PR4**	Sphingosine 1-phosphate receptor 4	19:3172343-3180329	-2.07	5.00E-05	0.0052
139410	**SDSL**	Serine dehydratase-like	12:113830249-113876081	-2.76	1.50E-03	0.012
75213	**SEMA3A**	Semaphorin-3A	7:83587658-84122040	2.11	0.00035	0.0219
74803	**SLC12A1**	Solute carrier family 12 member 1	15:48481233-48596275	-3.29	5.00E-05	0.0052
177542	**SLC25A22**	Mitochondrial glutamate carrier 1	11:790474-798316	-2.23	5.00E-05	0.0052
183751	**TBL3**	Transducin beta-like protein 3	16:2022037-2034193	-2.10	1.00E-04	0.0355
163931	**TKT**	Transketolase	3:53258722-53290068	-2.94	5.00E-05	0.0052
184281	**TSSC4**	Protein TSSC4	11:2421717-2425106	-2.29	1.50E-03	0.012
263563	**UBBP4**	Protein UBBP4	17:21729600-21731762	-2.05	5.00E-05	0.0052
160446	**ZDHHC12**	Probable palmitoyltransferase ZDHHC12	9:131464801-131486406	-2.37	5.00E-05	0.0052

After the filtering process, only the genes which displayed the largest change in expression (log2 fold change >2 and <-2) at a high significance (q<0.05) were chosen for further analysis (N=60). The genes are organised alphabetically.

3.4.2.2 Gene ontology

The list of significantly differentially regulated genes (N=60; Table 3.11) was sorted into gene ontology (GO) term categories for molecular function and biological processes using the programs DAVID (https://david.ncifcrf.gov/) and PANTHER (http://pantherdb.org/). The molecular function GO terms associated with the differentially regulated genes include a large percentage annotated as “binding activity” (39.7%) and “catalytic activity” (34.5%) (Figure 3.8). Other differentially regulated gene molecular function GO terms included those for “receptor activity” (5.2%); “structural molecule activity” (6.9%); “transporter activity” (12.1%) and “transcription factor activity” (1.7%). The biological process of differentially regulated genes GO terms included “metabolic process” (26.9%), “cellular process” (25.8%), “developmental process” (9.0%); “immune system process” (9.0%); “localization” (6.7%); “biological regulation” (5.6%); “biogenesis” (4.5%); “biological adhesion” (3.4%) and “apoptotic process” (1.2%) (Figure 3.8).

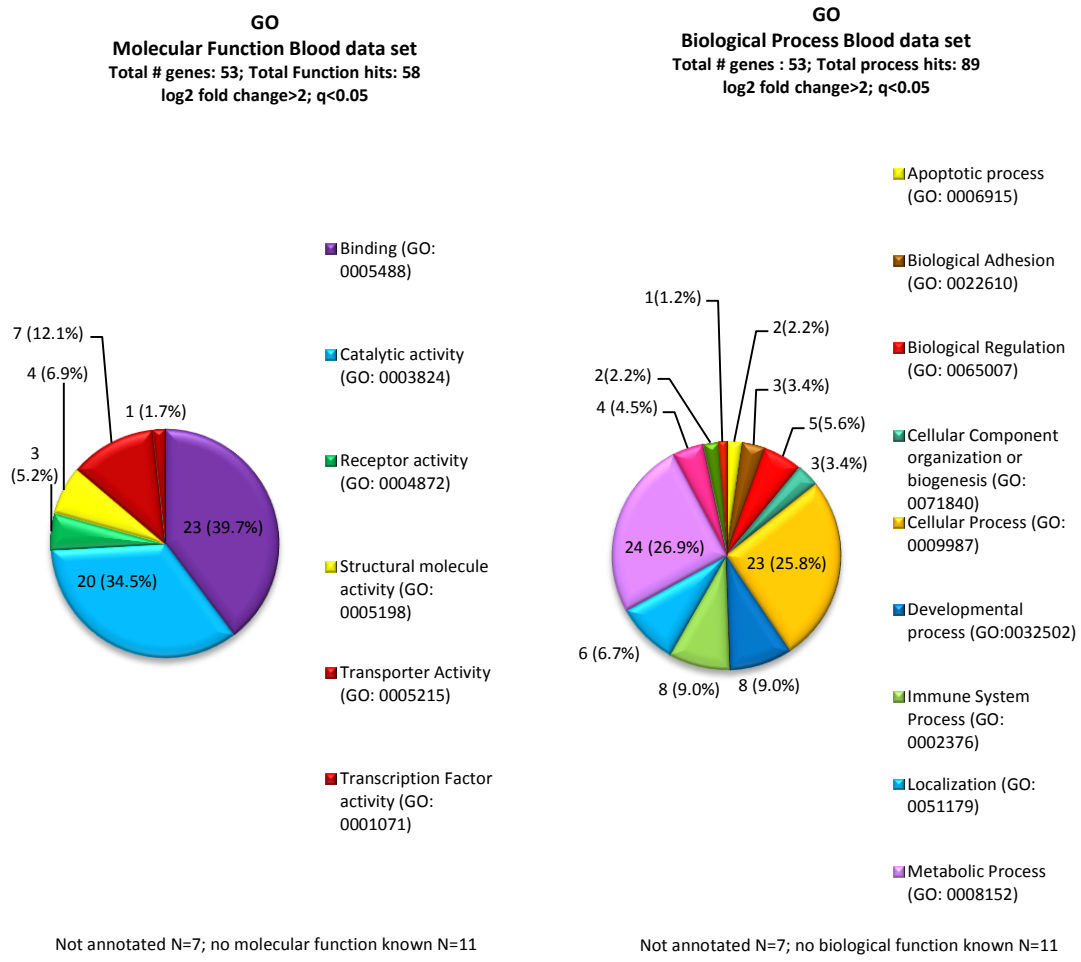


Figure 3.8: Functional categories based on the molecular process and biological function for the top 60 genes using DAVID and PANTHER.

Table 3.12 lists these top 60 genes with their molecular function and biological process. The genes which function in metabolic processes are shown in bold in red. From these gene ontology analyses, metabolic processes were among the most significant processes represented by the gene set. Diabetes is a metabolic disorder, and for this reason, these genes are strong functional candidates. These genes include; *AKT2 (*RAC-beta serine/threonine-protein kinase); *ALDOA (*Fructose-bisphosphate aldolase A) ; *ATP5D (*ATP synthase subunit delta, mitochondrial); *B3GNT8 (*BetaGalbeta-1,3,N acetylglucosaminyltransferase8); *DCXR (*Xylulose reductase); *G6PD (*Glucose-6-phosphate

dehydrogenase); *GAA (*Lysosomal alpha glucosidase); *GSK3A (*Glycogen synthase kinase-3 alpha); *HIST1H2BO (*Histone H2B type 1-O); *ISG15 (*Ubiquitin-like protein ISG15); *NAPRT1 (*nicotinate phosphoribosyltransferases); *NCF1C (*Putative neutrophil cytosol factor 1C); *PDZD3* (Na(+)/H(+) exchange regulatory cofactor NHE-RF4); *PGLS (*6-phosphogluconolactonase); *PTGDS (*Prostaglandin-H2 D-isomerase); *PYCR1* (Pyrroline-5-carboxylate reductase 1, mitochondrial), *TKT* (Transketolase) and *UBBP4* (Protein UBBP4).

Table 3.12: **Gene ontology results using DAVID and PANTHER for the top 60 genes found to be differentially expressed in the blood data set**

Gene_id ENSG000000	Gene	Molecular Function	Biological Process
002726	**ABP1**	Binding	Cellular
105221	**AKT2**	Binding; Catalytic	Apoptotic; Biological regulation; Biogenesis; Cellular; Developmental; Metabolic
149925	**ALDOA**	Binding; Catalytic	Metabolic
099624	**ATP5D**	Catalytic; Transporter	Cellular; localization; Metabolic
175756	**AURKAIP1**	No hit	Developmental
177191	**B3GNT8**	Catalytic	Cellular; Developmental; Metabolic; Multicellular organismal; Reproductive
168062	**BATF2**	Binding; Transcription Factor	Cellular; biological regulation
185347	**C14orf80**	No hit	No hit
173369	**C1QB**	Binding; Catalytic	Immune system response
007080	**CCDC124**	No hit	No hit
229119	**CTB-63M22,1**	Not annotated	No hit
169738	**DCXR**	Catalytic	Metabolic
164741	**DLC1**	Catalytic	Developmental
147647	**DPYS**	Catalytic	Metabolic
228502	**EEF1A1P11**	Not annotated	No hit
225663	**FAM195B**	Catalytic; Receptor	Apoptotic; Biological regulation; Cellular; Developmental
160211	**G6PD**	Catalytic	Metabolic
171298	**GAA**	Catalytic	Metabolic
169704	**GP9**	No hit	Biological adhesion; Cellular
250510	**GPR162**	Not annotated	No hit
105723	**GSK3A**	Catalytic	Cellular; Developmental; Metabolic; multicellular organismal
188536	**HBA2**	Binding; Transporter	Cellular
086506	**HBQ1**	Binding; Transporter	Localization; Multicellular organismal
196331	**HIST1H2BO**	Binding	Biogenesis; Cellular;

203813	**HIST1H3H**	Not annotated	No hit
146678	**IGFBP-1**	Binding	Biological Regulation, Cellular; Response to stimuli; Metabolic
211895	**IGHA1**	Binding	Immune system response
211890	**IGHA2**	Binding	Immune system response
211677	**IGLC2**	Binding	Immune system response
211662	**IGLV3-21**	Binding	Immune system response
185507	**IRF7**	No hit	Immune system process; response to stimulus
187608	**ISG15**	Binding; Structural	Metabolic
156113	**KCNMA1**	Transporter	Biological regulation; Cellular; localization
053918	**KCNQ1**	Transporter	Biological regulation; Cellular; Localization
110811	**LEPREL2**	Binding; Catalytic	Biological adhesion; Cellular
112139	**MDGA1**	No hit	Developmental
162576	**MXRA8**	No hit	No hit
147813	**NAPRT1**	Catalytic	Metabolic
165178	**NCF1C**	Binding; Catalytic	Cellular
174886	**NDUFA11**	No hit	Cellular
107281	**NPDC1**	No hit	No hit
079156	**OSBPL6**	No hit	Cellular; Localization
172367	**PDZD3**	Catalytic	Metabolic; Cellular
130313	**PGLS**	Catalytic	Metabolic
257704	**PRR24**	Not annotated	No hit
107317	**PTGDS**	Binding; Catalytic	Cellular;; localization
183010	**PYCR1**	Catalytic	Metabolic; Cellular
265150	**RN7SL2**	Not annotated	No hit
236552	**RPL13AP5**	Structural	Cellular
215030	**RPL13P12**	Not annotated	No hit
125910	**S1PR4**	Receptor	Cellular
139410	**SDSL**	Binding	Biological adhesion
075213	**SEMA3A**	Binding	Cellular; Developmental; Immune system; Multicellular organismal
074803	**SLC12A1**	Transporter	Cellular
177542	**SLC25A22**	Binding; Structural; Transporter	Localization; Metabolic
183751	**TBL3**	Binding	Biogenesis; Cellular; Metabolic
163931	**TKT**	Catalytic	Metabolic
184281	**TSSC4**	No hit	No hit
263563	**UBBP4**	Binding; Structural	Metabolic
160446	**ZDHHC12**	Binding	No hit

- No hit: No known recorded biological process or molecular function in PANTHER and/or DAVID database

Genes that have not yet been annotated (*CTB63M22*,1; *EEF1A1P11*; *GPR162*; *HIST1H3H*; *PRR44*; *RN7SL2* and *RPL13P12*) were excluded from further analysis. In order to identify what pathways these genes may play a role in, the PANTHER and KEGG (http://www.kegg.jp/kegg/pathway.html) gene list classification systems were used. Table 3.13 lists the known pathways that these genes are associated with. Only 11 genes out of the 60 genes of interest had known links to annotated pathways in the KEGG database. Pathways of interest include the fructose galactose metabolism pathway (P02744); glycolysis (P00024); insulin/IGF pathway (P00033) and the pentose phosphate pathway (P02762) as they all have a function in the metabolism of sugars (in bold in red in Table 3.13). The genes associated with these pathways are *ALDOA*; *AKT2*; *GSK3A*; *TKT*; *DCXR; PGLS*; and *G6PD*.

Table 3.13: **Pathway analysis using PANTHER and KEGG**

PANTHER/KEGG pathway	Gene
Axon guidance (P000007)	SEMA3A, AKT2
Angiogenesis (P00005)	AKT2
Apoptosis Signalling pathway (P00006)	AKT2
Blood coagulation (P00011)	GP9
Endothelin signalling pathway (P00019)	AKT2
FGF signalling pathway (P00021)	AKT2
Fructose galactose metabolism (P02744)	**ALDOA**
Glycolysis (P00024)	**ALDOA**
Heterotrimeric G-protein signalling pathway (P00026)	GSK3A
Huntington disease (P00029)	AKT2
Hypoxia response via HIF activation (P00030)	AKT2
Inflammation mediated by chemokine and cytokine signalling pathway (P00031)	AKT2
Insulin/IGF pathway (P00033)	**GSK3A, AKT2**
Interleukin signalling pathway (P00036)	AKT2
Oxidative phosphorylation (66043)	ATP5D
PDGF signalling pathway (P00047)	GSK3A; AKT2
PI5 kinase pathway (P00048)	AKT2
Pentose phosphate pathway (P02762)	**ALDOA, TKT, DCXR, G6PD, PGLS**
Proline biosynthesis (P02768)	PYCR1
Ras pathway (P04393)	GSK3A
RIG-I-like receptor signalling pathway	ISG15
T cell activation (P00053)	AKT2
p53 pathway (P00059)	AKT2

- Pathways in bold in red indicate pathways of interest as they all function to some extent in the metabolism of sugars
- FGF: Fibroblast Growth Factor; IGF: Insulin Growth Factor; PDGF: Platelet derived growth factor; RIG: retinoic acid-inducible gene 1

3.4.2.3 Visual inspection of gene expression within biological groups

To further reduce and prioritize the list of potential candidate genes, the gene expression pattern was examined in each individual sample (as opposed to gene expression within a biological group). A gene with a similar expression profile among cases and controls or that showed considerable variation within a biological group was excluded from further analysis. Genes with an expression pattern observed to be similar in all six of the GDM women and similar in all six controls, but showing significant differences between cases when compared to controls were selected for further analysis. *MDGA1* was initially one of the top candidates because it had a statistically significant (qvalue<0.05) lower level of expression (-4.75 fold) (under-expression) in the case group (Figure 3.9 A). However, when analysing the expression levels of this gene in each individual case and control (Figure 3.9 B); it became clear that control individuals 0, 1, 3 and 5 have similar levels of *MDGA1* expression seen in all the case individuals. For this reason, the lower expression of *MDGA1* is unlikely to play a role in the aetiology of gestational diabetes as low expression is common in control individuals. This gene was excluded from further analysis.

Figure 3.9 C shows that there is a substantially lower expression of *G6PD* in women with GDM when compared to women with normal glucose tolerance during pregnancy. When investigating the expression of this gene in each individual, it becomes clear that in each case, there is less gene expression when compared to controls, who have evidently higher expression of *G6PD* (Figure 3.9 D). For this reason, *G6PD* was included for further evaluation as a potential candidate in the development of gestational diabetes. Student t-tests were performed to confirm the results of the visual inspection to ensure that only genes that display significant differential expression (log2 fold change) between cases and controls, are included in further analysis.

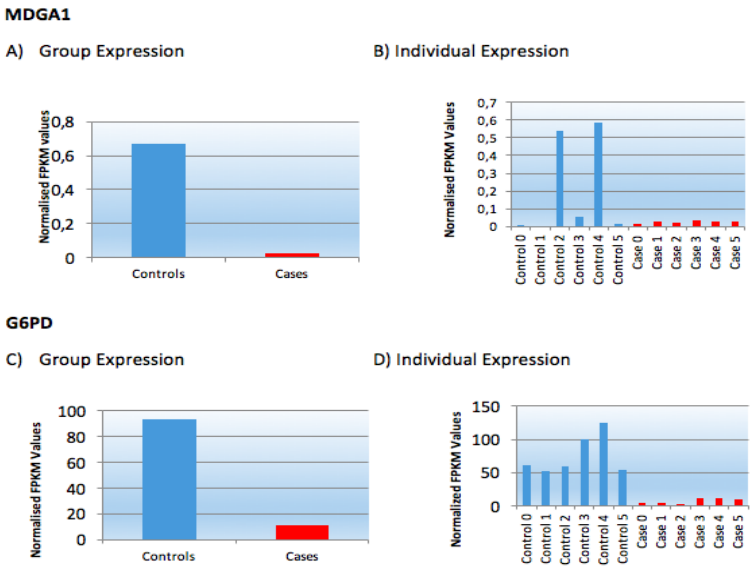


Figure 3.9: **Visual inspection of the normalised average gene expression of each gene in in each group (case and control) (A and C).** B and D shows the normalised expression of the gene in each individual sample.

Fifteen of the 60 genes displayed significant differential expression unique to cases when compared to controls and were included for further analysis. Gene enrichment analysis was done using these 15 genes as an input. These genes include ***AKT2*** *(*V-akt murine thymoma viral oncogene homolog 2*);* ***ALDOA*** *(*Aldolase A, Fructose Biphosphate*);* ***C14orf80****;* ***DCXR*** *(*Dicarbonyl/L-Xylulose reductase*);* ***G6PD*** *(*Glucose-6-phosphate dehydrogenase*);* ***GSK3A*** *(*Glycogen Synthase Kinase 3-Alpha*);* ***KCNQ1*** *(*potassium channel, voltage gated KQT-like subfamily Q, member 1*);* ***PGLS*** *(*6-phosphogluconolactonase*);* ***SLC25A22*** (Solute carrier Family 25/Mitochondrial carrier, glutamate member 22); ***TKT*** *(*Transketolase*);* ***UBBP4*** ***(***ubiquitin B pseudogene 4*);* ***SEMA3A*** (Semaphorin 3A); ***GAA*** *(acid alpha glucosidase);* ***PDZD3*** (PDZ domain containing 3) and ***ISG15*** (Ubiquitin-like protein ISG15).

3.4.2.4 Gene Enrichment

Gene enrichment (functional enrichment analysis) is a method to identify classes of genes or proteins that are over-represented in a large set of genes or proteins. The method uses statistical approaches to identify significantly enriched or depleted groups of genes. Table 3.14 shows the outcome of the gene enrichment for the above mentioned 15 genes. The

gene enrichment output highlighted a group of genes (N=5) that were significantly enriched for a particular GO term. These five genes (*G6PD*, *TKT*, *ALDOA*, *PGLS* and *DCXR*) are significantly associated together with the GO terms "pentose metabolic process"; "NADP metabolic process" and "pyridine nucleotide metabolic process". Upon further investigation, it was observed that these 5 genes encode enzymes that function in the pentose phosphate pathway (Figure 3.10).

After doing an extensive literature search using each of the 15 genes as subjects and taking into consideration the gene enrichment and gene ontology output, the list of potential candidates was reduced to ten strong functional and biological candidates. Table 3.15 lists these ten genes and provides further rationale for their selection.

Table 3.14: Gene enrichment using PANTHER

	GO Biological Process	# genes from ref list (Homo Sapiens)	# genes from Input list	Fold Enrichment	p Value
A	Pentose Metabolic process (GO: 0005996)	12	**3**	>5	2.1E-02
B	NADP Metabolic Process (GO: 0006734)	14	**4**	>5	4.8E-03
C	Pyridine Nucleotide Metabolic Process (GO: 0019362)	92	**5**	>5	7.8E-04

	Gene	Gene Name	log2 fold change	q-Value
1	DCXR	Xylulose reductase	-2.18	0.030
2	G6PD	Gluscose-6-phosphate dehydrogenase	-3.09	0.005
3	PGLS	6-Phosphogluconolactonase	-2.52	0.020
4	ALDOA	Fructose-bisphosphate aldolase A	-2.26	0.009
5	TKT	Transketolase	-2.93	0.005

Figure 3.10: The pentose phosphate pathway illustrating the genes found to be differentially expressed between GDM cases and controls.

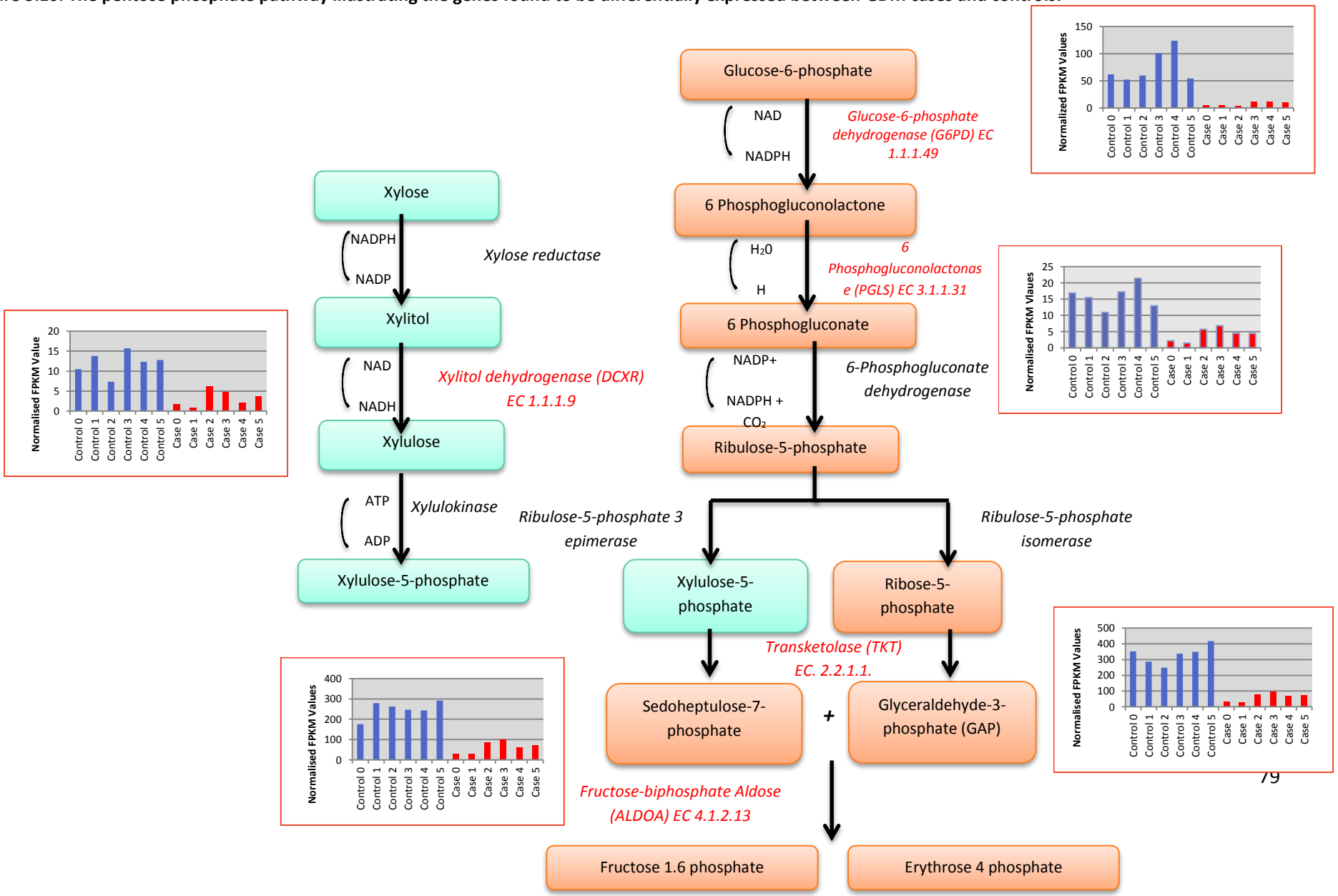
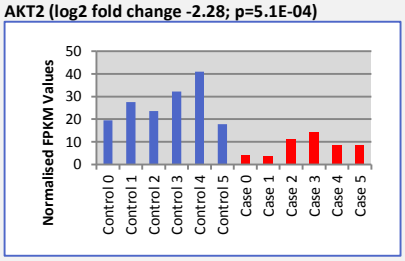
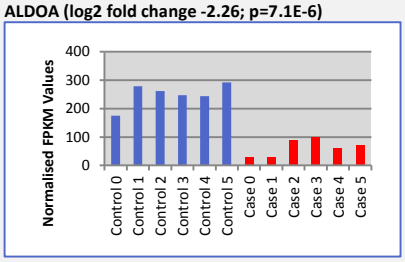
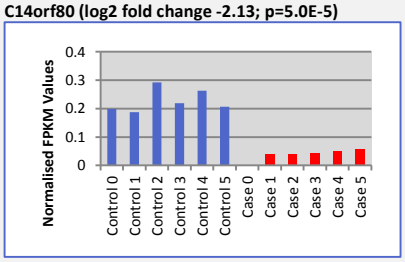
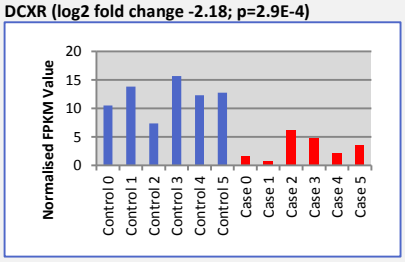
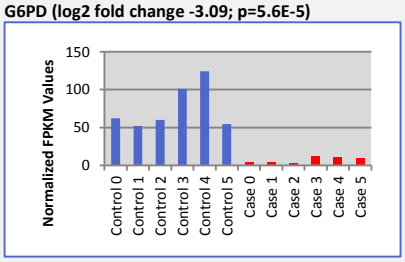
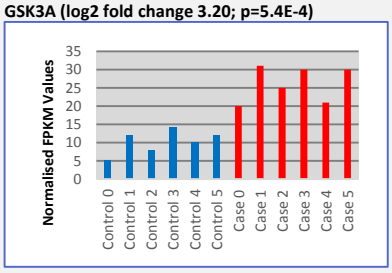
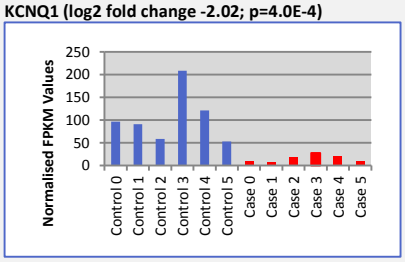
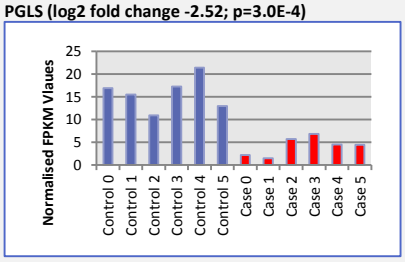
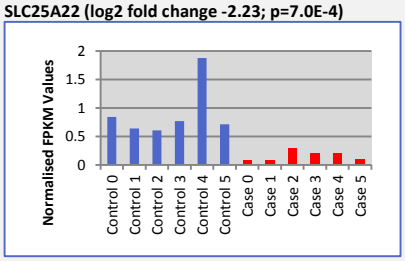
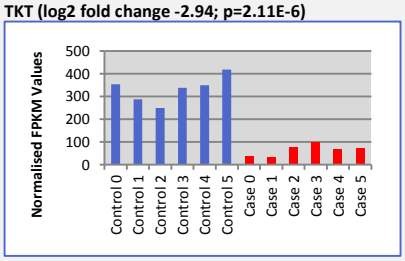


Table 3.15: The top ten genes chosen as candidates for validation for differential gene expression in the blood data set (p-values were calculated using a Students t-test)

Gene	Gene Name	Reason for validation
AKT2 (log2 fold change -2.28; p=5.1E-04) 	RAC-beta serine/threonine-protein kinase	**AKT2 has been implicated in T2D** * A Family with Severe Insulin Resistance and Diabetes Mellitus due to a Missense Mutation in AKT2 * Contribution of the **AKT2** gene to type 2 diabetes in the Chinese Han population: The **AKT2** protein kinase is thought to be a key mediator of the insulin signal transduction process, **AKT2** is suggested to play a role in glucose metabolism and the development or maintenance of proper adipose tissue and islet mass * Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta): show that mice deficient in Akt2 are impaired in the ability of insulin to lower blood glucose because of defects in the action of the hormone on liver and skeletal muscle, These data establish **AKT2** as an essential gene in the maintenance of normal glucose homeostasis (Cho et al., 2001, George et al., 2004, Sun et al., 2011).
ALDOA (log2 fold change -2.26; p=7.1E-6) 	Fructose-bisphosphate aldolase A	Key enzyme in the **Pentose Phosphate Pathway** and catalyses the transformation between dihydroxyacetone phosphate, glyceraldehyde-3- phosphate and fructose 1, 6 - biphosphate in the glycolytic pathway, * ALDOA was found to be down regulated in brown adipose tissue, which might indicate impaired glucose utilization (Medrikova et al., 2015, Meugnier et al., 2007).
C14orf80 (log2 fold change -2.13; p=5.0E-5) 	Uncharacterized protein C14orf80	Possible **marker** for uncontrolled gestational diabetes
DCXR (log2 fold change -2.18; p=2.9E-4) 	Xylulose reductase	The protein encoded by this gene acts as a homotetramer to catalyze diacetyl reductase and L-xylulose reductase reactions, The encoded protein may play a role in the uronate cycle of **glucose metabolism** and in the cellular osmoregulation in the proximal renal tubules, Mutations in this gene are a cause of pentosuria ((Lee et al., 2013).
G6PD (log2 fold change -3.09; p=5.6E-5) 	Gluscose-6-phosphate dehydrogenase	G6PD is the rate-limiting enzyme of the **pentose phosphate pathway** * Time-resolved metabolomics analysis of β-cells implicates the pentose phosphate pathway in the control of insulin release pathway, Pathophysiologic roles for G6PD have also been identified in such disease processes as diabetes, aldosterone-induced endothelial dysfunction, cancer, and others, * Effects of laparoscopic Roux-en-Y gastric bypass on glucose-6 phosphate dehydrogenase activity in obese **type 2 diabetic**s: G6PD overexpression has been implicated in insulin resistance, hyperlipidaemia, and increased oxidative stress in animals (Schneider et al., 2012, Spegel et al., 2013)

GSK3A (log2 fold change 3.20; p=5.4E-4) 	Glycogen synthase kinase-3 alpha	GSK3A is a key target for the **development of novel treatments for T2D**, GSK-3 has recently been the subject of much research because it has been implicated in a number of diseases, including Type II diabetes (Diabetes mellitus type 2) * Analysis of hepatic gene transcription in mice expressing insulin-insensitive * GSK3 regulation is proposed to play a key role in the hormonal control of many cellular processes, Inhibition of GSK3 in animal models of diabetes leads to normalization of blood glucose levels, while high GSK3 activity has been reported in Type II diabetes (Gokhale and Tilak, 2013; Lipina et al,, 2015)
KCNQ1 (log2 fold change -2.02; p=4.0E-4) 	Potassium voltage-gated channel subfam KQT member	KCNQ1 has been identified as a **susceptibility gene for T2D** * Genome-wide association studies in Japanese and Dutch populations recently identified common variants in the KCNQ1 gene to be associated with type 2 diabetes (Been et al., 2011, Kasuga, 2011)
PGLS (log2 fold change -2.52; p=3.0E-4) 	6-phosphogluconolactonase	Gene found in the **Pentose Phosphate Pathway** No other information in the literature linking this gee with glucose metabolism.
SLC25A22 (log2 fold change -2.23; p=7.0E-4) 	Mitochondrial glutamate carrier 1	MiR-184 regulates insulin secretion through repression of SLC25A22: report that miR-184 **inhibits insulin secretion** in the MIN6 pancreatic cell line through the r**epression of its target Slc25a22**, a mitochondrial glutamate carrier, The study provides new insight into the regulation of insulin secretion by glutamate transport in mitochondria ((Morita et al., 2013).
TKT (log2 fold change -2.94; p=2.11E-6) 	Transketolase	Gene found in **pentose phosphate pathway** * Effect of high dose thiamine therapy on activity and molecular aspects of transketolase in Type 2 diabetic patients: All enrolled Type 2 diabetics had > 40% lower mononuclear transketolase activity as compared to healthy individuals, * Genetic variability in enzymes of metabolic pathways conferring protection against non-enzymatic glycation versus diabetes-related morbidity and mortality; hypothesized that genetic variability in genes encoding enzymes metabolizing glycolytic intermediates produced in excess under hyperglycemic conditions [i,e,, transketolase (TKT), transaldolase, TKT-like protein 1, fructosamine 3-kinase (FN3K), glyoxalase 1 and glucose-6-phosphate dehydrogenase] could influence progression of diabetic nephropathy (DN) and diabetes-related morbidity and mortality (Halim et al., 2013, Tanhauserova et al., 2014).

3.5 Analysing data generated from PLACENTA (exposed vs unexposed) comparisons

3.5.1 A simple filtering process to reduce the number of possible candidate genes

The data were first sorted to only include genes with an "OK" status (N=19880). Only a few genes had the LOWDATA (N=26) and HIDATA (N=6) outcome and were subsequently excluded. The second filtering step removed all the genes that did not display significant differential expression (qvalue>0.05) between the cases and controls. This filtering step reduced the number of potential candidates substancially. The total number of genes displaying highly significant (qvalue<0.05) differential expression beween the two groups was 1088. The third filtering step focused on genes with significant (qvalue<0.05) differential expression greater than 1 fold between the two biological groups. This gene list was further reduced by focussing on a genes which had a differential expression greater than 2 fold between the two biological groups. These genes were targeted as potential candidates as they displayed the largest difference in expression between GDM cases and controls. This reduced the number of genes of interest to 52 (Figure 3.11).

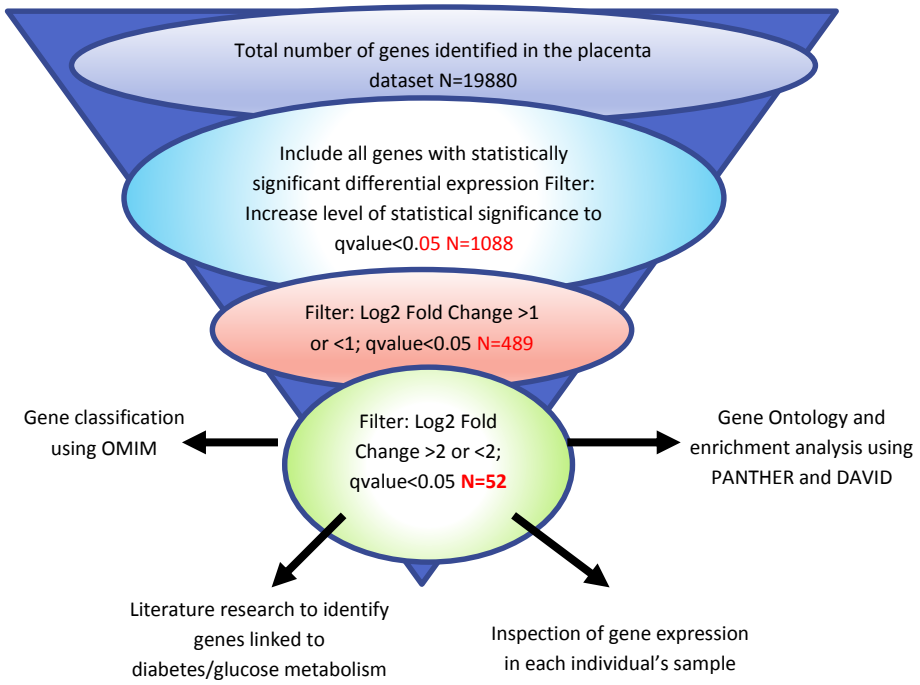


Figure 3.11: **Flow diagram depicting how potential candidate genes from the placenta data set were selected for validation.**

3.5.1.1 The identification of genes displaying significant differential expression between cases and controls

The volcano plot (Figure 3.12) plots the relationship between the p-values of a statistical test and the magnitude of the difference in expression of 19880 genes. The $-\log_{10}$ p-values are plotted on the y-axis and the log2 fold change values are plotted on the x-axis. The red circles indicate the genes of interest that display both large-magnitude fold-changes (x-axis) as well as high statistical significance (-log10 of p-value, y-axis). The dashed red-line shows where q = 0.05 with points above the line having qvalue < 0.05 and points below the line having qvalue > 0.05. This plot is colored such that the green points represent the genes with significant (qvalue>0.05) fold-change of more than 1 (log2 < 1); the blue dots represent genes with a significant log2 fold change of >2 or <-2. The red dots represent genes with the largest significant (qvalue<0.05) differential expression (log2 fold change >3 or <-3) (log2 >2).

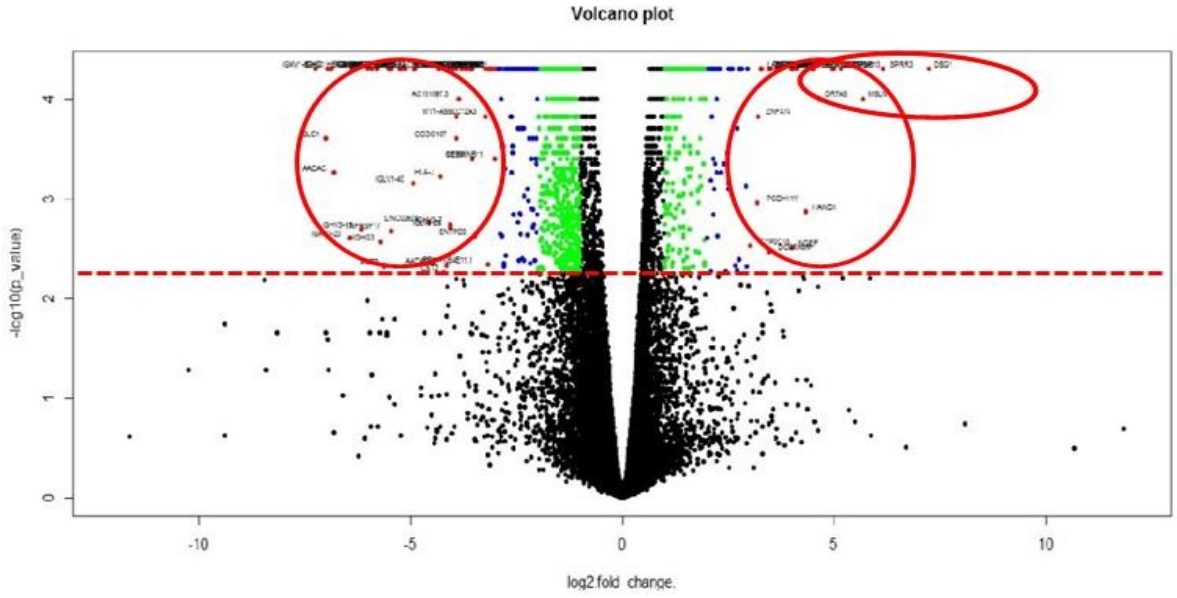


Figure 3.12: The Volcano Plot illustrating significant fold change differential expression for placenta data set comparing cases to controls. Volcano plot for the 19880 genes from the placenta data set (log2 fold change >1; qvalue<0.05: green; log2 fold change >2; qvalue<0.05: blue; log2 fold change >3; qvalue<0.05: red).

The scatter plot (Figure 3.13) illustrates the relationship between the corrected p-value (q-value) and the log2 fold change value of the genes showing the largest change in expression and largest significance (red dots from the volcano plot).

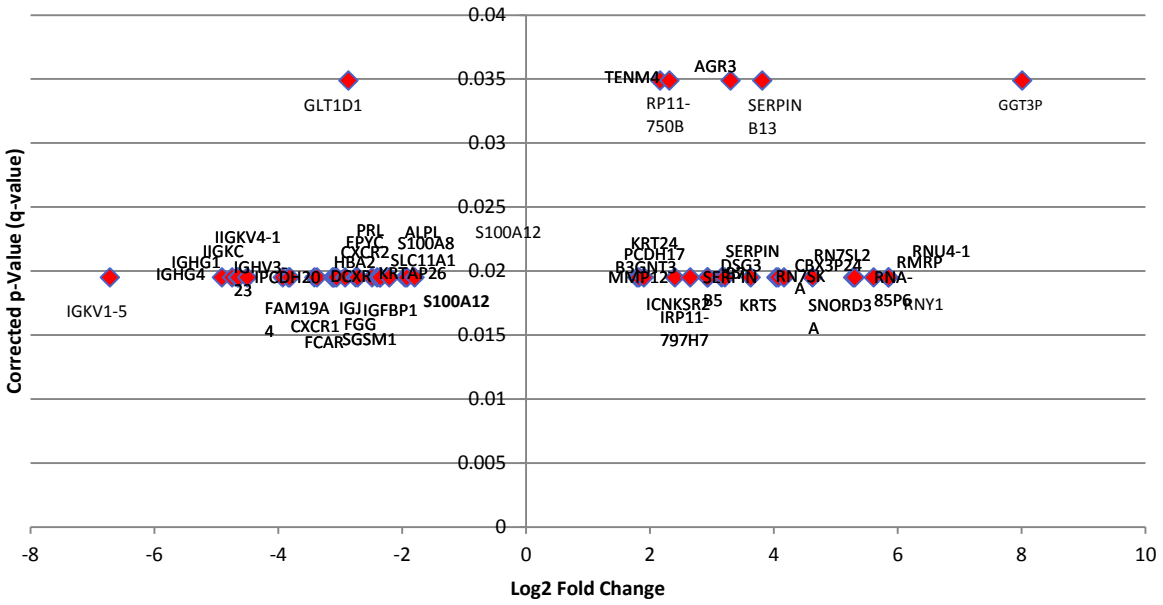


Figure 3.13: Scatter Plot illustrating significant fold change differential expression for placenta data set comparing cases to controls. The scatter plot shows only the genes which have a log2 fold change >2 and <-2; qvalue<0.05 (corrected for multiple testing).

These 52 genes (shown in the scatter plot) were used for further analysis. Table 3.16 lists the gene names, their location within the genome, the log2 fold change values of each gene, the corresponding p-value (showing level of significance) as well as the qvalue. The qvalues are the Benjamini-Hochberg corrected p-values for multiple-testing.

Table 3.16: **List of genes displaying significant differential expression between cases and controls from placental tissue (log2 fold change >2 and >-2; qvalue < 0.05)**

gene_id ENSG00000	gene	Name	locus	Log2 fold change	p_value	q_value
173467	AGR3	Anterior gradient protein 3 homolog	7:16899028-16921611	3.31	1.00E-04	0.0349
162551	ALPL	Alkaline phosphatase. tissue-nonspecific isozyme	1:21835857-21904905	-2.49	5.00E-05	0.0195
179913	B3GNT3	UDP-GlcNAc:betaGal beta-1.3-N-acetylglucosaminyltransferase 3	19:17905636-17923891	2.08	5.00E-05	0.0195
266405	CBX3P2	Not annotated	18:2652168-2655394	4.07	5.00E-05	0.0195
149970	CNKSR2	Connector enhancer of kinase suppressor of ras 2	X:21392535-21672813	2.66	5.00E-05	0.0195
250182	CTD-2165H16.1	Not annotated	5:14652046-14653438	-3.84	2.00E-04	0.0195
163464	CXCR1	C-X-C chemokine receptor type 1	2:219027567-219031718	-3.37	5.00E-05	0.0195
180871	CXCR2	C-X-C chemokine receptor type 2	2:218990011-219001976	-2.35	1.00E-04	0.0349
169738	DCXR	L-xylulose reductase	17:79993011-79995608	-2.40	5.00E-05	0.0195
134757	DSG3	Desmoglein-3	18:29027757-29058665	3.16	5.00E-05	0.0195
83782	EPYC	Epiphycan	12:91357455-91398803	-2.72	4.60E-03	0.0195
163377	FAM19A4	family with sequence similarity 19 (chemokine (C-C motif)-like). member A4	3:68780916-68981761	-3.42	5.00E-05	0.0195
186431	FCAR	Immunoglobulin alpha Fc receptor	19:55385548-55401838	-3.12	5.00E-05	0.0195
160211	G6PD	Glucose-6-phosphate dehydrogenase	X:153759605-153796782	-2.68	5.00E-05	0.0195
197421	GGT3P	Putative gamma-glutamyltranspeptidase 3	22:18761177-18830912	8.01	1.00E-04	0.0349
151948	GLT1D1	Glycosyltransferase 1 domain-containing protein 1	12:129337971-129469509	-2.87	1.00E-04	0.0349
146678	IGFBP-1	Insulin-like growth factor-binding protein 1	7: 45888357-45893668	-4.74	3.00E-05	0.0195
115457	IGFBP-2	Insulin-like growth factor-binding protein 2	2: 216632828 - 216664436	-2.65	4.00E-05	0.0195
211896	IGFBP-6	Insulin-like growth factor-binding protein 6	7:45927955-45933267	-2.92	5.00E-05	0.0195
211892	IGHG4	immunoglobulin heavy constant gamma 4 (G4m marker)	14:106090686-106092403	-4.91	5.00E-05	0.0195
211949	IGHV3-23	immunoglobulin heavy variable 3-23	14:106725200-106725733	-4.5	4.00E-05	0.0195
132465	IGJ	Immunoglobulin J chain	4:71494460-71552533	-3.11	5.00E-05	0.0195
211592	IGKC	similar to hCG26659; immunoglobulin kappa constant	2:89109983-89165653	-4.65	3.00E-05	0.0195
243466	IGKV1-5	Ig kappa chain V-I region HK102 (Fragment)	2:89246818-89247475	-6.72	5.00E-05	0.0195
241351	IGKV3-11	Not annotated	2:89326667-89327228	-3.81	5.00E-05	0.0195
211598	IGKV4-1	Ig kappa chain V-IV region (Fragment)	2:89184912-89185669	-4.64	5.00E-05	0.0195

167916	KRT24	Keratin. type I cytoskeletal 24	17:38854242-38860002	2.41	1.00E-05	0.0349
186081	KRT5	Keratin. type II cytoskeletal 5	12:52908358-52914471	3.63	5.00E-05	0.0195
197683	KRTAP26-1	Keratin-associated protein 26-1	21:31691451-31692607	-2.05	5.00E-05	0.0195
110347	MMP12	Macrophage metalloelastase	11:102733466-102745764	2.03	5.00E-05	0.0195
118946	PCDH17	Protocadherin-17	13:58205943-58303445	2.12	5.00E-05	0.0195
197991	PCDH20	Protocadherin-20	13:61983990-62002220	-3.93	5.00E-05	0.0195
172179	PRL	Prolactin	6:22260652-22318027	-2.75	3.00E-05	0.0195
199916	RMRP	RNA component of mitochondrial RNA processing endoribonuclease	9:35657750-35658014	5.31	5.00E-05	0.0195
202198	RN7SK	Not annotated	6:52860417-52860748	4.04	5.00E-05	0.0195
265150	RN7SL2	Not annotated	14:50329270-50329567	4.17	5.00E-05	0.0195
251705	RNA5-8SP6	Not annotated	Y:10037763-10037915	5.28	5.00E-05	0.0195
200795	RNU4-1	Not annotated	12:120730899-120731040	5.61	5.00E-05	0.0195
202538	RNU4-2	Not annotated	12:120729565-120729706	5.85	5.00E-05	0.0195
201098	RNY1	Not annotated	7:148684227-148684340	5.86	5.00E-05	0.0195
262902	RP11-750B16.1	Not annotated	17:51183094-51183719	2.32	0.0001	0.0349
234338	RP11-797H7.1	Not annotated	7:64295657-64297260	2.93	5.00E-05	0.0195
163221	S100A12	Protein S100-A12	1:153346183-153348125	-2.21	5.00E-05	0.0195
143546	S100A8	Protein S100-A8	1:153362507-153363664	-2.01	5.00E-05	0.0195
197641	SERPINB13	Serpin B13	18:61254222-61271873	3.82	0.0001	0.0349
206075	SERPINB5	Serpin B5	18:61143993-61172318	2.94	5.00E-05	0.0195
166396	SERPINB7	Serpin B7	18:61420168-61472604	3.21	5.00E-05	0.0195
167037	SGSM1	Small G protein signalling modulator 1	22:25202235-25323545	-3.06	5.00E-05	0.0195
18280	SLC11A1	Natural resistance-associated macrophage protein 1	2:219246751-219261617	-2.04	5.00E-05	0.0195
20236	SNORD3A	Not annotated	17:19091328-19092027	4.63	5.00E-05	0.0195
163931	TKT	Transketolase	3: 53224707 - 53256114	-2.56	0.0001	0.0349

After the filtering process, only the genes which displayed the largest change in expression (log2 fold change >2 and >-2) at the highest statistical significance (qvalue<0.05) were chosen for further analysis (N=52).

3.5.1.2 Gene ontology

The list of significantly differentially regulated genes (N=52) was sorted into gene ontology (GO) term categories for molecular function and biological processes using the programs DAVID (https://david.ncifcrf.gov/) and PANTHER (http://pantherdb.org/). The molecular function GO terms associated with the differentially regulated genes include a large percentage of "binding activity" (39.1%) and "catalytic activity" (26.1%). Other differentially regulated gene molecular function GO terms included those for "receptor activity" (13%); "enzyme regulator activity" (10.9%); "structural molecule activity" (6.4%); "transporter activity" (4.3%) (Figure 3.14). Biological process of differentially regulated genes GO terms included "immune system process" (20.8%); "metabolic process" (16.7%); "cellular process" (12.5%); "developmental process" (12.5%); "biological adhesion" (4.2%); "biological regulation" (9.4%); "developmental process" (11.5%) and "localization" (12.5%) (Figure 3.14).

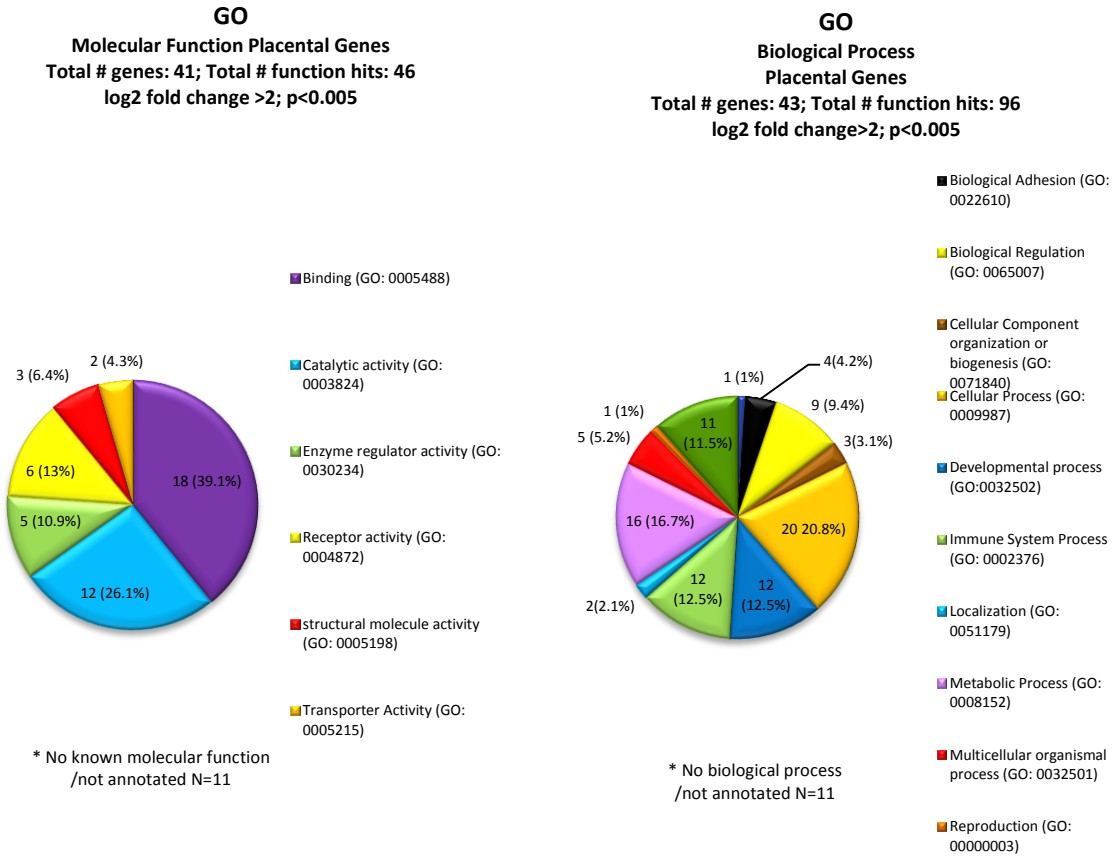


Figure 3.14: The molecular function and biological process of the significantly differentially expressed genes identified from the placenta data set.

Table 3.17 lists these top 52 genes with their molecular functions and biological processes. The genes which play a role in metabolic processes are shown in bold in red. From this gene ontology analysis, metabolic and immune system processes were among the most significant processes represented by the gene set. Diabetes is a metabolic disorder, and for this reason, these genes are strong functional candidates. These genes include; ***CNKSR2*** *(*Connector enhancer of kinase suppressor of ras 2); ***CXCR1*** (C-X-C chemokine receptor type 1); ***CXCR2*** (C-X-C chemokine receptor type 2) ***EPYC*** (Epiphycan); ***G6PD*** (Glucose-6-phosphate dehydrogenase); ***GLT1D1*** (Glycosyltransferase 1 domain-containing protein 1); ***GGT3P*** (Putative gamma-glutamyltranspeptidase 3); ***IGFBP-1*** (Insulin growth factor binding 1), ***IGFBP-2*** (Insulin growth factor binding 2); ***IGFBP-6*** (Insulin growth factor binding 6); ***MMP12*** (Macrophage metalloelastase); ***S100A12*** (Protein S100-A12); ***S100A8*** (Protein S100-A8); ***SerpinB5***; ***SerpinB7***; ***SerpinB13***; ***SGSM1*** (Small G protein signalling modulator 1) and ***TKT*** (Transketolase)**.** Only 4 of these genes are associated with known pathways (Table 3.18).

Table 3.17: Gene ontology using DAVID and PANTHER for genes found to be differentially expressed in placenta samples from cases when compared to controls (log2 fold change>2; qvalue<0.05)

Gene_id (ENSG00000)	Gene	Molecular function	Biological Process
162551	**ALPL**	Binding	Developmental
266405	**CBX3P2**	No hit	No hit
149970	**CNKSR2**	Binding; Catalytic; Enzyme Regulator	Biological Regulation; **Metabolic**
250182	**CTD-2165H16,1**	No hit	No hit
163464	**CXCR1**	Catalytic; Receptor	Cellular; Immune System; **Metabolic**; Response to stimulus
180871	**CXCR2**	Catalytic; Receptor	Cellular; Immune System; **Metabolic**; Response to stimulus
134757	**DSG3**	Binding	Cellular Process
83782	**EPYC**	Receptor	Biological Adhesion; Cellular; Immune System; **Metabolic**
163377	**FAM19A4**	No hit	No hit
186431	**FCAR**	Binding; Receptor	Cellular; Immune System; Response to stimulus
171557	**FGG**	Binding	Biological Adhesion; Cellular; Response to stimulus
197421	**G6PD**	Catalytic	**Metabolic**
197421	**GGT3P**	Catalytic	**Metabolic**
151948	**GLT1D1**	Catalytic	**Metabolic**
188536	**HBA2**	Binding	Cellular

146678	**IGFBP-1**	Binding	Biological Regulation; Cellular; Response to stimulus; **metabolic**
211896	**IGFBP-2**	Binding	Biological Regulation; Cellular; Response to stimulus; **metabolic**
211892	**IGFBP-6**	Binding	Biological Regulation; Cellular; Response to stimulus; **metabolic**
211949	**IGHV3-23**	Binding	Immune System
132465	**IGJ**	Binding	Immune System; Response to stimulus
211592	**IGKC**	Binding	Immune System
243466	**IGKV1-5**	Binding	Developmental; Immune System; Response to stimulus
241351	**IGKV3-11**	Binding	Immune System
211598	**IGKV4-1**	Binding	Developmental; Immune System; Response to stimulus
167916	**KRT24**	Structural Molecule	Biogenesis; Cellular Process; Developmental
186081	**KRT5**	Structural Molecule	Biogenesis; Cellular; Developmental
197683	**KRTAP26-1**	Structural Molecule	Cellular Process
110347	**MMP12**	Binding; Catalytic	Cellular Process, Metabolic
118946	**PCDH17**	Binding	Biological Adhesion; Cellular; Developmental
197991	**PCDH20**	Binding	Cellular Process
172179	**PRL**	Binding	Cellular; Developmental
199916	**RMRP**	No hit	No hit
202198	**RN7SK**	No hit	No hit
265150	**RN7SL2**	No hit	No hit
251705	**RNA5-8SP6**	No hit	No hit
200795	**RNU4-1**	No hit	No hit
202538	**RNU4-2**	No hit	No hit
201098	**RNY1**	No hit	No hit
262902	**RP11-750B16,1**	No hit	No hit
234338	**RP11-797H7,1**	No hit	No hit
163221	**S100A12**	Binding	Cellular; Immune System; **Metabolic**; Response to stimulus
143546	**S100A8**	Binding	Cellular; **Metabolic**
197641	**SERPINB13**	Catalytic; Enzyme Regulator	Biological Regulation; **Metabolic**
206075	SERPINB5	Catalytic; Enzyme Regulator	Biological Regulation; **Metabolic**
166396	SERPINB7	Catalytic; Enzyme Regulator	Biological Regulation; **Metabolic**
167037	SGSM1	Binding; Catalytic; Enzyme Regulator	Biological Regulation,; Cellular; Developmental; Localization; **Metabolic**
18280	SLC11A1	Transporter	Localization; Response to stimulus

202,364,263,934	SNORD3A	No hit	No hit
149256	TKT	Catalytic	**Metabolic**

Table 3.18: **Pathway analysis using PANTHER and KEGG**

PANTHER/KEGG pathway	Gene
Inflammation mediated by chemokin and cytokine signalling pathway (P00031)	CXCR1; CXCR2
Pentose phosphate pathway (P02762)	**TKT, G6PD**
p53 pathway (P00059)	SerpinB5
PI3 Kinase Pathway (P00048)	IGFBP-1

3.5.1.3 Visual inspection of gene expression

To further reduce the list of potential candidate genes, the expression pattern of each gene was analysed in each individual sample (as opposed to an overall expression in a group). Genes which displayed similar levels of varied expression within cases and within controls where the differences were not consistent between cases compared to controls were excluded from further analysis. Genes with an expression pattern unique to cases were included for further analysis. *CBX3P2* was also one of the top potential candidates because it had statistically significant (P=5.0E-05) differential expression (log2 fold change) of 4.07. However, when analysing this gene in the case versus controls as a group (Figure 3.15 A) it is evident that the expression of *CBX3P2* is substantially higher in the case group than in the controls. However, when looking at the expression of this gene in each individual (Figure 3.15 B); it becomes clear that case 2 and case 4 have similar expression levels of *CBX3P2* to control individuals. For this reason, it is unlikely that the over-expression of *CBX3P2* contributes to the aetiology of gestational diabetes and is therefore excluded from further analysis.

Figure 3.15 C illustrates the substantially lower expression of *IGFBP-1* in women with GDM when compared to women with normal glucose tolerance during pregnancy. When investigating the expression of this gene in each individual, it becomes clear that in each case, *IGFBP-1* is significantly under-expressed when compared to controls, who have evidently higher expression of *IFGBP-1* (Figure 3.15 D). For this reason, *IGFBP-1* was included for further

evaluation as a potential candidate in the development of gestational diabetes. Student t-tests were performed to confirm the results of the visual inspection to ensure that only genes that display significant differential expression (log2 fold change) between cases and controls, were included in further analysis.

Twenty (N=20) of the 52 genes displayed significant differential expression unique to the cases when compared to the controls. These twenty genes are ***CXCR1*** (C-X-C chemokine receptor type 1); ***CXCR2*** (C-X-C chemokine receptor type 2); ***DSG3*** (Desmoglein-3); ***G6PD*** (Glucose-6-phosphate dehydrogenase); ***GGT3P*** (Putative gamma-glutamyltranspeptidase 3); ***GLT1D1*** (Glycosyltransferase 1 domain-containing protein 1); ***IGFBP-1*** (Insulin-like growth factor-binding protein 1); ***IGFBP-2*** (Insulin-like growth factor-binding protein 2); ***IGFBP-6*** (Insulin-like growth factor-binding protein 6); ***IGKV1*-5** (Ig kappa chain V-I region HK102 (Fragment); ***PCDH2*** (Protocadherin-20); ***RMRP*** (RNA component of mitochondrial RNA processing endoribonuclease); ***RN7SK*** (Not annotated); ***RN7SL2*** (Not annotated); ***S100A12*** (Protein S100-A12); ***SNORD3A*** (Not annotated); ***TKT*** (Transketolase); ***SGSM2*** (Small G protein signalling modulator 2); ***SLC11A1*** (Natural resistance-associated macrophage protein 1) and ***TKT*** (Transketolase). These twenty genes were used as the input for a gene enrichment analysis using PANTHER.

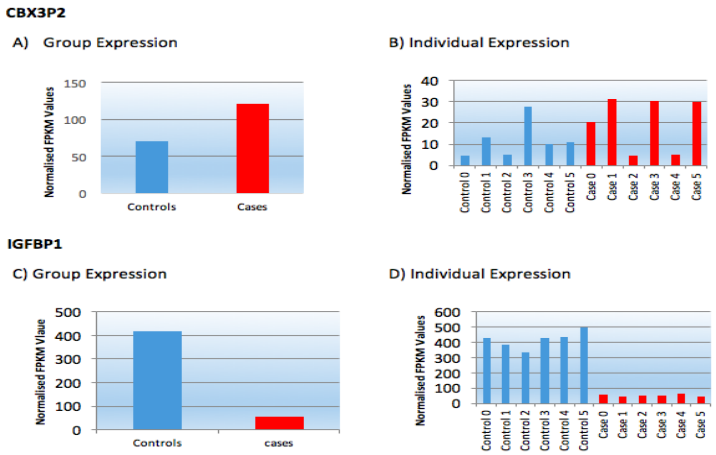


Figure 3.15: **Visual inspection of the normalised gene expression in the case versus control groups (A, C) as well as in each individual sample (B, D).**

3.5.1.4 Gene Enrichment

The gene enrichment output highlighted two genes, *CXCR1* and *CXCR2* that were significantly enriched for the particular GO terms "interleukin-8-mediated signalling pathway" (GO:0038112); "cellular response to interleukin-8" (GO:0098759); and "response to interleukin-8" (GO:0098758). There have been studies which associate an increase in interleukin-8 with diabetes (Dakovic et al., 2013, Srinivasan et al., 2004, Zozulinska et al., 1999)(Table 3.19). The process of gene enrichment also highlight *TKT* and *G6PD* to be significantly associated with the GO term "pentose metabolic process". These two genes were also found to be significantly under expressed in the blood of women with GDM. These genes transcribe key enzymes in the PPP.

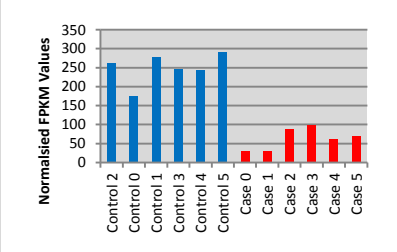
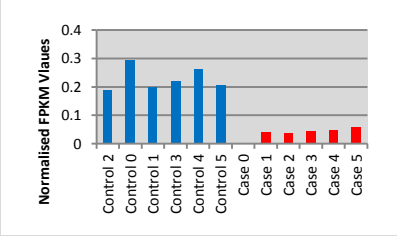
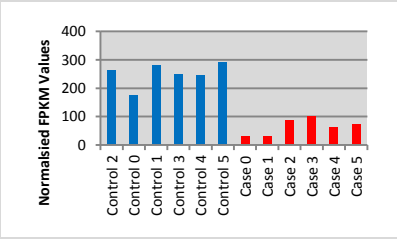
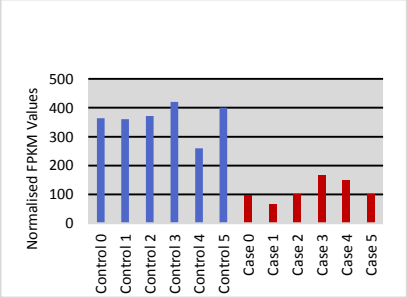
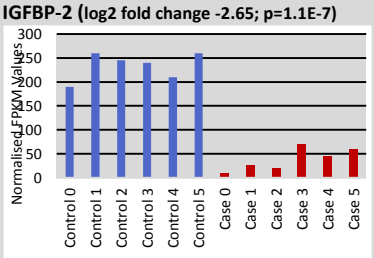
Table 3.19: Gene enrichment of placental dataset using PANTHER

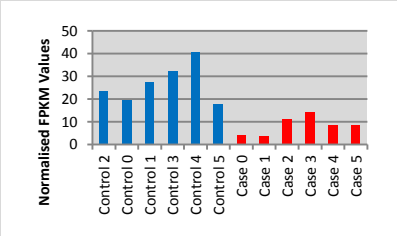
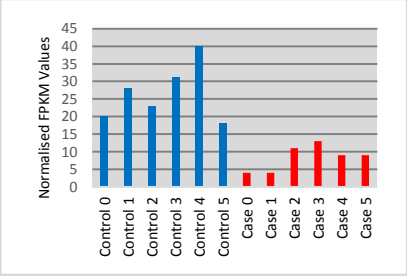
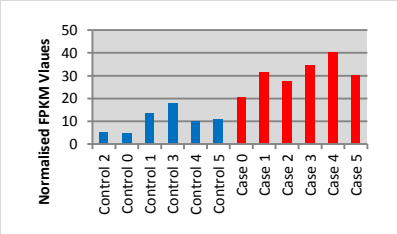
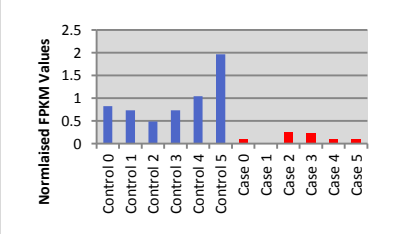
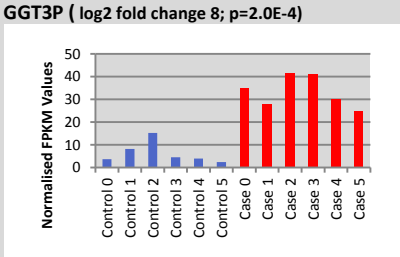
	GO Biological Process	# genes from ref list (Homo Sapiens)	# genes from Input list	Fold Enrichment	p-value
A	Interleukin-8-mediated signalling pathway (GO:0038112)	2	2	>5	0.042
B	Cellular response to Interleukin-8 (GO:0098759)	2	2	>5	0.045
C	Response to interleukin-8 (GO:0098758)	2	2	>5	0.056
D	Pentose metabolic process (GO: 0005996)	2	2	>5	0.003

	Gene	Gene Name	log2 fold change	q-Value
1	CXCR1	C-X-C chemokine receptor type 1	-3.84	0.019
2	CXCR2	C-X-C chemokine receptor type 2	-3.37	0.029
3	G6PD	Glucose-6-phosphate dehydrogenase		
4	TKT	Transketolase		

After doing an extensive literature search for each of these top 20 genes and taking into consideration the gene enrichment and gene ontology output, the list was reduced to ten genes believed to be strong functional and biological candidates. Table 3.20 lists these ten genes and provides a rationale for their selection.

Table 3.20: The top ten genes chosen as candidates for validation for differential gene expression

Gene	Gene Name	Reason to validate
CXCR1 (log2 fold change -3.37; p=4.6E-6) 	Chemokine Receptor 1	* CXCR1/CXCR2 pathway - involvement in diabetes pathophysiology * CXCR1/2 Inhibition blocks and reverts T1D in mice * Type 1 Diabetes Prone NOD Mice Have Diminished Cxcr1 mRNA Expression in Polymorphonuclear Neutrophils and CD4+ T Lymphocytes, Although numerous chemokine/chemokine receptor pathways have been described to be implicated in the pathogenesis of type 1 diabetes (T1D), the CXCR1/2 axis has recently been proved to be crucial for leucocyte recruitment involved in insulitis and β cell damage (Citro et al., 2015a, Citro et al., 2015b, Haurogne et al., 2015).
CXCR2 (log2 fold change -2.35; p=5.0E-5) 	Chemokine Receptor 2	* CXCR1/CXCR2 pathway - involvement in diabetes pathophysiology * CXCR1/2 Inhibition blocks and reverts T1D in mice * Type 1 Diabetes Prone NOD Mice Have Diminished Cxcr1 mRNA Expression in Polymorphonuclear Neutrophils and CD4+ T Lymphocytes, Although numerous chemokine/chemokine receptor pathways have been described to be implicated in the pathogenesis of type 1 diabetes (T1D), the CXCR1/2 axis has recently been proved to be crucial for leucocyte recruitment involved in insulitis and β cell damage (Citro et al., 2015a, Citro et al., 2015b, Haurogne et al., 2015)
G6PD (log2 fold change -2.68; p=2.9E-6) 	Glucose-6-phosphate dehydrogenase	G6PD is the rate-limiting enzyme of the pentose phosphate pathway * Time-resolved metabolomics analysis of β-cells implicates the pentose phosphate pathway in the control of insulin release pathway, Pathophysiologic roles for G6PD have also been identified in such disease processes as diabetes, aldosterone-induced endothelial dysfunction, cancer, and others, * Effects of laparoscopic Roux-en-Y gastric bypass on glucose-6 phosphate dehydrogenase activity in obese type 2 diabetics: G6PD overexpression has been implicated in insulin resistance, hyperlipidaemia, and increased oxidative stress in animals (Schneider et al., 2012, Spegel et al., 2013).
IGFBP-1 (log2 fold change -4.74; p=8.3E-6) 	Insulin like growth factor binding protein 1	Increased DNA methylation levels of IGFBP1- are associated with T2D in Swedish men * Low concentrations of IGFBP-1 are associated with insulin resistance, diabetes and cardiovascular disease, Insulin-like growth factor-binding protein 1 and 7 concentrations are lower in obese pregnant women, women with gestational diabetes and their foetuses (Lappas, 2015). Inverse changes in foetal insulin-like growth factor (IGF)-1 and IGF binding protein-1 in association with higher birth weight in maternal diabetes (Lindsay et al., 2007). Insulin-like growth factor axis and gestational diabetes: A longitudinal study in a multiracial cohort (Zhu et al., 2016).
IGFBP-2 (log2 fold change -2.65; p=1.1E-7) 	Insulin like growth factor binding protein 2	Insulin-like growth factors (IGFs) and IGF-binding proteins (*IGFBP-1, -2* and *-3)* in diabetic pregnancy: relationship to macrosomia (Yang et al., 1996). Insulin-like growth factor axis and gestational diabetes: A longitudinal study in a multiracial cohort (Zhu et al., 2016).

IGFBP-6 (log2 fold change -3,37; p=7.1E-4) 	Insulin like growth factor binding protein 6	* Expression of IGFBP-6 increased in patients with T1D, IGFBP-6 is an O-linked glycoprotein that preferentially binds IGF-II, inhibiting IGF-II actions including proliferation, survival and differentiation of a wide range of cells, IGFBP-6 levels were higher in patients with type 1 diabetes and its complications, although there was substantial overlap with control subjects (Bach, 2015c, Bach, 2015b, Lu et al., 2012).)Insulin-like growth factor axis and gestational diabetes: A longitudinal study in a multiracial cohort (Zhu et al., 2016).
TKT (log2 fold change -2.56; p=5.1E-5) 	Transketolase	Gene found in pentose phosphate pathway * Effect of high dose thiamine therapy on activity and molecular aspects of transketolase in Type 2 diabetic patients: All enrolled Type 2 diabetics had > 40% lower mononuclear transketolase activity as compared to healthy individuals, * Genetic variability in enzymes of metabolic pathways conferring protection against non-enzymatic glycation versus diabetes-related morbidity and mortality; hypothesized that genetic variability in genes encoding enzymes metabolizing glycolytic intermediates produced in excess under hyperglycemic conditions [i,e,, transketolase (TKT), transaldolase, TKT-like protein 1, fructosamine 3-kinase (FN3K), glyoxalase 1 and glucose-6-phosphate dehydrogenase] could influence progression of diabetic nephropathy (DN) and diabetes-related morbidity and mortality ((Halim et al., 2013, Tanhauserova et al., 2014).
MMP12 (log2 fold change 2.03; p=5.0E-3) 	MMP12 matrix metallopeptidase 12	* Elevated MMP12 levels are associated with atherosclerotic burden and symptomatic cardiovascular disease in subjects with T2D, The plasma level of MMP-7 and -12 were found to be elevated in type 2 diabetes mellitus and higher levels were associated with more severe atherosclerosis and an increased incidence of coronary events (Goncalves et al., 2015).
GLT1D1 (log2 fold change -2.87; p=1.0E-4) 	Glycosyltransferase 1 domain containing 1	No known association with diabetes but had almost 3 fold lower expression cases (log 2 fold change = 2.87)
GGT3P (log2 fold change 8; p=2.0E-4) 	Putative gamma-glutamyltranspeptidase 3	No known association with diabetes but had a much higher expression in cases (log 2 fold change = 8.01)

3.6 RNA-seq differential expression validation using single gene TaqMan assays

The validation of the RNA-seq results was limited to the ten most likely candidate genes for the blood dataset and ten genes for the placenta dataset (twenty genes were validated in total). This subset of genes was chosen using specific criteria including the magnitude of differential gene expression (fold change), the significance levels observed (p-value) and also their role in pathways linked to glucose metabolism. All of the TaqMan assays were successfully optimised and permitted data analysis for the purpose of validation. The TaqMan methodology is the gold-standard for validation of RNA-seq experiments. The validation using TaqMan showed complete concordance of expression with the RNA-seq results. Before the data for the target genes was analysed, the data from the three housekeeping genes was analysed to determine that they had stable and consistent gene expression across the samples. Figure 3.16 illustrates the C_T values of the three housekeeping genes for each blood (A) and placenta sample (B) (average of the triplicates). The C_T values are relatively consistent and show little variation among the samples regardless of disease status. This indicates stable expression which is desirable for a housekeeping gene. For this reason, all three housekeeping genes were used for normalization of target expression in the samples.

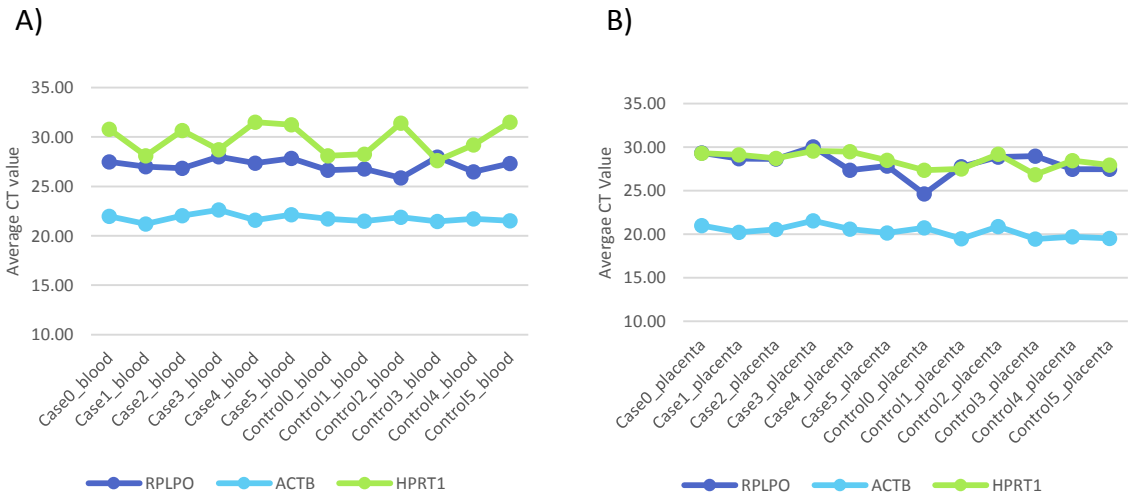


Figure 3.16: The C_T values across all samples (cases and controls) for the three housekeeping genes for A) the blood samples and B) the placental samples. These graphs indicate that all three genes have stable and consistent gene expression cross the samples and can be used for the normalization.

Figures 3.17 and 3.19 illustrate the normalised expression values of each individual sample generated via the TaqMan method. The expression observed in the cases was significantly altered compared to that observed in the controls and this was similar to the patterns of expression observed using RNA-seq. Figures 3.18 and 3.20 show that the RNA-seq data had a linear relationship with qRT–PCR (a goodness of fit (R^2) of 0.937 for the blood samples and 0.923 for the placental samples). R^2 values of 0.9 - 1 indicate a strong correlation between the two techniques.

Table 3.21 qRT-PCR validation of RNA-seq data on a selection of ten potential candidate genes from the blood data set

	RNA-seq				qRT-PCR				
	Normalised FPKM Values				Normalised expression Value[2]			Fold Change	
Gene	Cases	Controls	Normalised fold changes[1]	p-value	$\triangle C_T$ Cases (SD)	$\triangle C_T$ Controls (SD)	$\triangle\triangle C_T$[3]	$2\text{-}\Delta\Delta C_T$[4]	p-value
ALDOA	63.50	301.98	-2.26	1.12E-04	-2.67 (0.14)	-1.70 (0.27)	-0.97	1.95	1.24E-05
G6PD	20.96	193.25	-3.20	5.01E-05	-5.07 (0.68)	-3.36 (0.16)	-1.17	2.25	7.27E-06
DCXR	3.15	28.33	-3.16	5.50E-04	-5.93 (1.47)	-5.10 (0.27)	-0.83	1.77	2.63E-04
PGLS	4.21	24.16	-2.52	3.10E-04	-3.04 (0.72)	-2.36 (0.17)	-0.68	1.60	1.86E-03
TKT	63.37	642.36	-3.36	5.00E-05	-4.35(0.67)	-3.02 (0.15)	-1.33	2.5	2.78E-05
C14ORF80	3.81	16.69	-2.13	1.00E-04	-2.34 (0.48)	-2.21. (0.41)	-0.13	1.09	2.54E-03
AKT2	8.46	40.91	-2.27	5.36E-04	-3.25 (0.59)	-2.4 (0.33)	-0.85	1.80	2.96E-05
GSK3A	146.97	10.21	3.84	5.00E-05	5.45(1.05)	6.95(0.10)	1.5	0.35	1.48E-03
KCNQ1	16.93	168.86	-3.32	1.12E-04	-5.09 (1)	-3.91 (0.39)	-1.18	2.26	4.83E-05
SLC25A22	15.51	72.96	-2.22	5.00E-05	-3.66 (0.41)	-2.30 (0.44)	-1.36	2.56	2.92E-07

[1]Normalized $\log_2$ fold changes: fold changes as $\log_2$ (normalized expression value control/normalized expression value case),
[2]ΔC_T cases = (Average C_T value Target gene - Average C_T value Housekeeping gene) case sample, ΔC_T controls = (Average C_T value Target gene - Average C_T value Housekeeping gene) control sample
[3] $\triangle\triangle C_T$ - ΔC_T cases - ΔC_T controls, Negative $\Delta\Delta C_T$ values indicate a lower expression of the target gene in the case group when compared to controls,
[4] $2^{-\Delta\Delta CT}$ is the fold change (not indicative of direction of expression only magnitude of expression)

A) B)

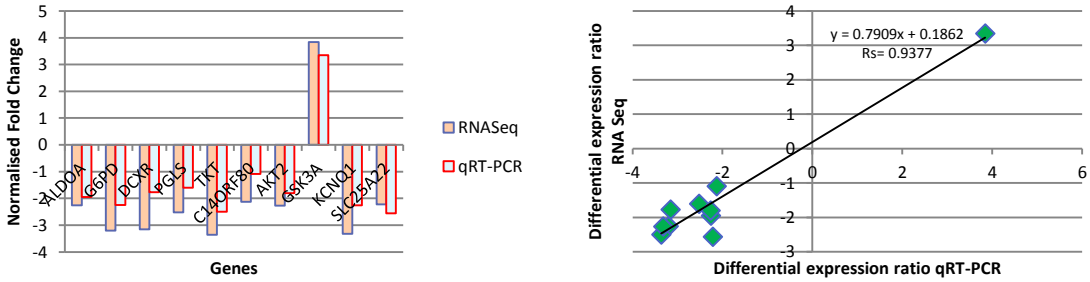


Figure 3.17: RT-qPCR validation of RNA-seq results. Ten genes from the blood data set were selected for differential expression confirmation in the same RNA samples used for RNA-seq. A: log 2 fold change comparison B: Correlation analysis between RNA-seq and RT-qPCR log 2 fold change results from the same RNA samples. Spearman correlation coefficient is displayed (Rs).

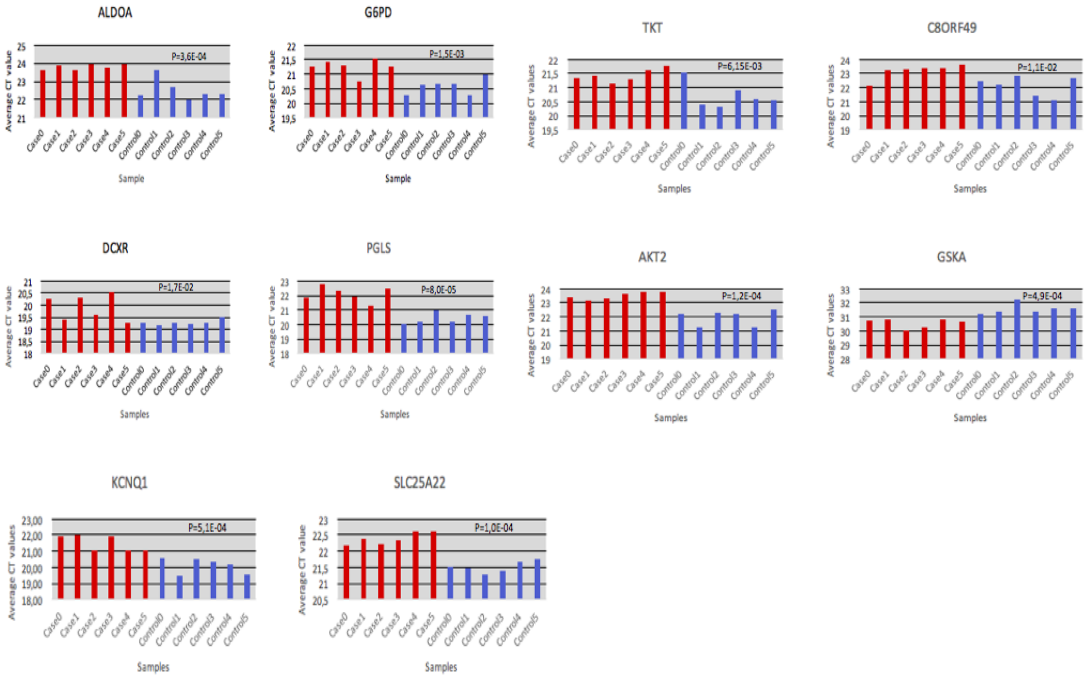


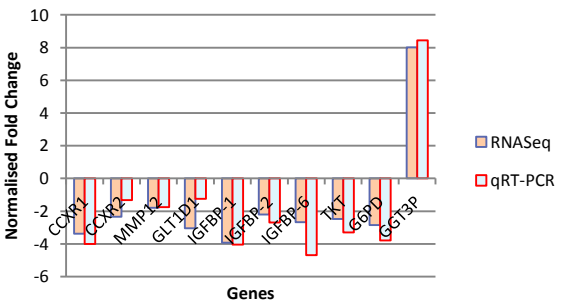
Figure 3.18. The average C_T values for each sample and for each of the ten genes chosen for validation from the blood data set. The C_T values for the case samples are shown in red and the control samples are shown in blue. Higher C_T values are indicative of a lower expression of mRNA.

Table 3.22 qRT-PCR validation of RNA-seq data on a selection of ten potential candidate genes from the placental dataset

	RNA-seq				qRT-PCR				
	Normalised FPKM Values				Normalised expression Value[2]				
Gene	Cases	Controls	Normalised fold changes[1]	p-value	ΔC_T Cases (SD)	ΔC_T Controls (SD)	$\triangle\triangle CT$[3]	$2\text{-}\Delta\Delta C_T$[4]	p-value
CXCR1	0.45	4.74	-3.37	5.00E-05	-6.00 (0.85)	-5.01 (0.68)	-4.0	4.00	3.6E-04
CXCR2	21.85	29.46	-2.34	5.00E-05	-9.36 (2.07)	-8.95 (0.58)	-0.41	1.33	1.4E-04
MMP12	52.87	15.22	1.79	1.00E-04	-8.68 (1.04)	-7.87 (1.71)	-0.81	1.75	3.2E-02
GLT1D1	0.14	1.23	-3.05	5.00E-05	-1.78 (0.80)	-1.46 (0.80)	-0.32	1.24	6.2E-02
IGFBP-1	55.40	419.66	-3.92	5.00E-05	-9.52 (2.3)	-7.50 (0.97)	-2.02	4.05	2.8E-02
IGFBP-2	6.52	30.13	-2.20	5.00E-05	-5.51 (3.3)	-4.42 (0.91)	-1.09	2.21	8.1E-04
IGFBP-6	20.65	86.43	-2.68	5.00E-05	-3.62 (3.96)	- 1.39 (0.71)	-2.23	4.69	7.36E-07
TKT	4.18	22.17	-2.47	4.23E-04	-6.18 (0.67)	-4.45 (0.62)	-1.73	3.31	1.6E-06
G6PD	0.16	1.19	-2.86	1.00E-04	-6.18 (0.67)	-4.25 (0.62)	-1.93	3.8	8.6E-04
GGT3P	0.30	0.0011	8.01	1.00E-04	6.55 (1.35)	9.63 (2.31)	-3.08	8.45	2.17E-07

[1]Normalized $\log_2$ fold changes: fold changes as $\log_2$ (normalized expression value control/normalized expression value case),
[2]ΔC_T cases = (Average C_T value Target gene - Average C_T value Housekeeping gene) case sample, ΔC_T controls = (Average C_T value Target gene - Average C_T value Housekeeping gene) control sample
[3] $\triangle\triangle C_T$ - ΔC_T cases - ΔC_T controls, Negative $\Delta\Delta C_T$ values indicate a lower expression of the target gene in the case group when compared to controls,
[4] $2^{-\Delta\Delta CT}$ is the fold change (not indicative of direction of expression only magnitude of expression)

A)



B)

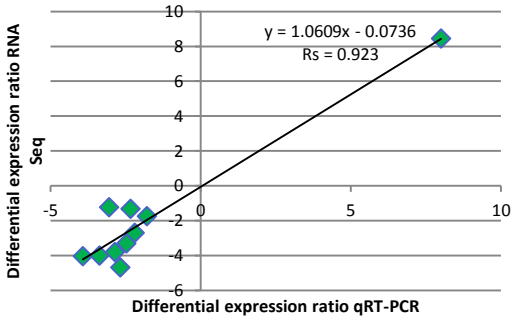


Figure 3.19: RT-qPCR validation of RNA-seq results. Ten genes from the placenta data set were selected for differential expression confirmation in the same RNA samples used for RNA-seq, A: log 2 fold change comparison B: Correlation analysis between RNA-seq and RT-qPCR log 2 fold change results from the same RNA samples. Spearman correlation coefficient is displayed (Rs).

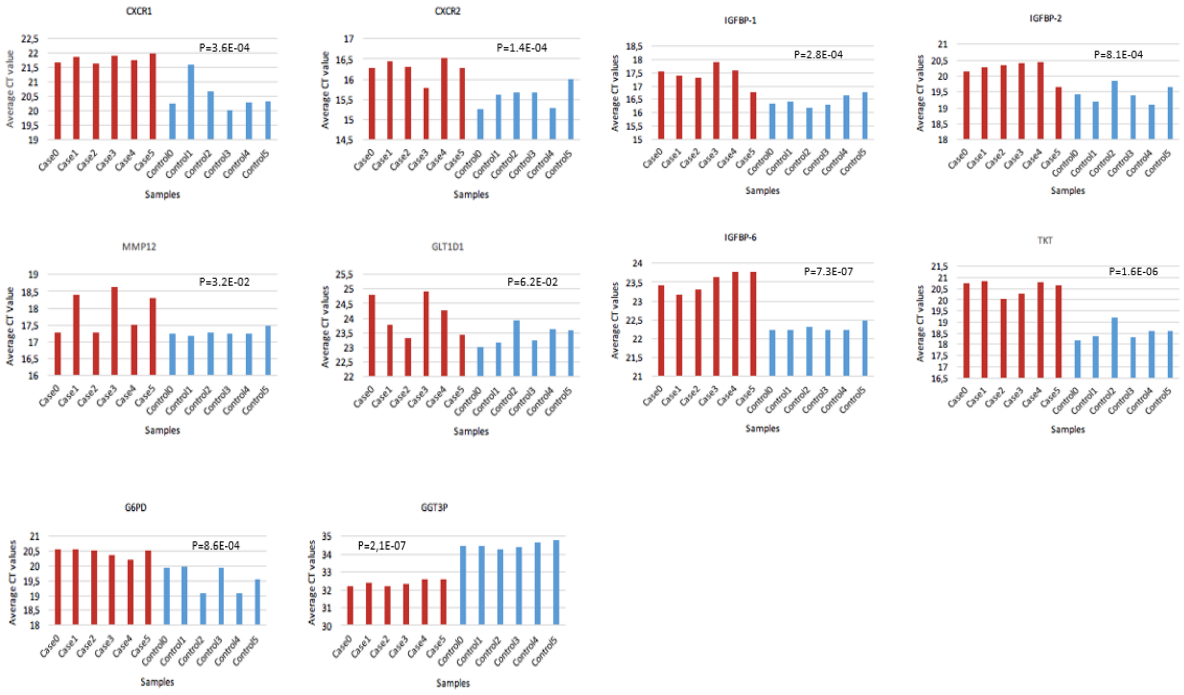


Figure 3.20. The average C_T values for each sample and for each of the ten genes chosen for validation from the placenta data set. The C_T values for the exposed samples are shown in red and the unexposed samples are shown in blue. Higher C_T values are indicative of a lower expression of mRNA.

From these twenty validated genes, only five could be selected for methylation analysis due to cost constraints. *G6PD* and *TKT* were selected based on the fact that they both display significant differential expression in the blood and placental samples and also encode enzymes that determine the rate at which the pentose phosphate pathway will function. The other three genes chosen for promoter region methylation analysis were the three insulin growth factor binding proteins *(IGFBP-1*, *IGFBP-2* and *IGFBP-6*) that were all significantly under expressed in placenta exposed to a GDM environment. These were considered important to analyse as one of the aims of this study was to determine epigenetic effects that may be observed in the placenta (as a proxy for the foetus) that had been exposed to the adverse *in utero* environment, in this case GDM.

3.7. Promoter region methylation analysis of selected genes (*G6PD*, *TKT*, *IGFBP-1*, *IGFBP-2* and *IGFBP-6*)

Methylation at a gene promoter region is a known epigenetic mechanism correlated with gene expression. Hypermethylation usually results in a decrease in gene expression whereas hypomethylation is correlated with increased gene expression. To explain the mechanism of altered gene expression of these genes, the methylation status of the promoter region was analyzed. Correlation between DNA methylation level and gene expression was investigated to assess to what degree gene expression may be influenced by altered DNA methylation in women who develop gestational diabetes and to determine whether this aberrant methylation was inherited by the foetus.

The C_T values of each of the digests (mock (M), methylation-sensitive (Ms), methylation-dependant (Md), double (Msd)) for each gene assay were in the range of what was expected (recommended)(Ceccarelli et al., 2016, Laska et al., 2013). The mock digests for all genes were within the range of 18 to 27 cycles and the C_T values of the Ms and Md digests were between the values of the mock and double digests. The C_T values of the double digests were higher than the C_T values of the mock digest (All the resulting C_T values are listed in Appendix J). The difference in C_T values between the double digest and mock digest samples represents the analytical window of the assay and should be greater than 3 (Laska et al., 2013). This means that more than 93.6 % of all DNA molecules in the samples were digested, and that the assay results are reliable and meaningful. For each gene assay, this criterion was met.

For the methylation-sensitive enzyme control (SEC), the difference in C_T values between the methylation-sensitive and mock digests should be equal to or greater than 4 to pass the quality control (Karatzas et al., 2014). Likewise for the methylation-dependent enzyme control (DEC), the difference in C_T values between the methylation dependent and mock digests should be equal to or greater than 4 to pass the quality control. (Agrogiannis et al., 2014, Duron et al., 2012, Gupta, 2015, Gupta et al., 2015)

3.7.1 DNA methylation analysis of candidate genes

Figure 3.21 shows the percentage (%) of DNA promoter region methylation observed for each gene analyzed in both the blood and placental samples and also lists the p-value which indicates whether the level of promoter region methylation between cases and controls reaches statistical significance. The p-value was calculated using a paired Student's t-test when comparing the blood values with the placental values and when comparing the cases with the controls, a non-paired student's t-test was used. There was no significant alteration in promoter region methylation for the IGFBPs (*IGFBP-1*, *IGFBP-2* and *IGFBP-6*) in the women who develop GDM when compared to the controls (blood samples) (p=0.85; p=0.91 and p=0.11 respectively; Figure 3.21 A, B and C). However, we do observe significant alterations in DNA methylation at the promoter region of these genes in the placenta. Thus, the promoter region of *IGFBP-1* and *IGFBP-2* in exposed placenta is significantly hypermethylated compared to the unexposed group ($p=2.2x10^{-8}$ and $p=2.x10^{-5}$, respectively) (Figure 3.21 A and B). For *IGFBP-6,* there appears to be a trend towards increased promoter region methylation in the placenta of the cases when compared to the controls, however this difference did not quite reach statistical significance (p=0.08, Figure 3.21 C).

The level of *G6PD* promoter region methylation in blood samples from cases was significantly higher ($p=1.9x10^{-5}$, Figure 3.21 D) than that seen in the controls. Similarly, the level of *G6PD* promoter region methylation in exposed placenta was significantly higher than that seen in unexposed placenta ($p=1.25x10^{-11}$, Figure 3.21D). The level of methylation at the *TKT* promoter region did not appear to be statistically significant between cases and controls in both the blood and placenta groups (Figure 3.21 E).

For the IGFBPs, the methylation levels are higher in the placenta than in the blood for both cases and control groups. This is not the case for *G6PD*, where the methylation levels are higher in the blood and placenta cases, and for *TKT*, where methylation is low across both tissues and biological groups.

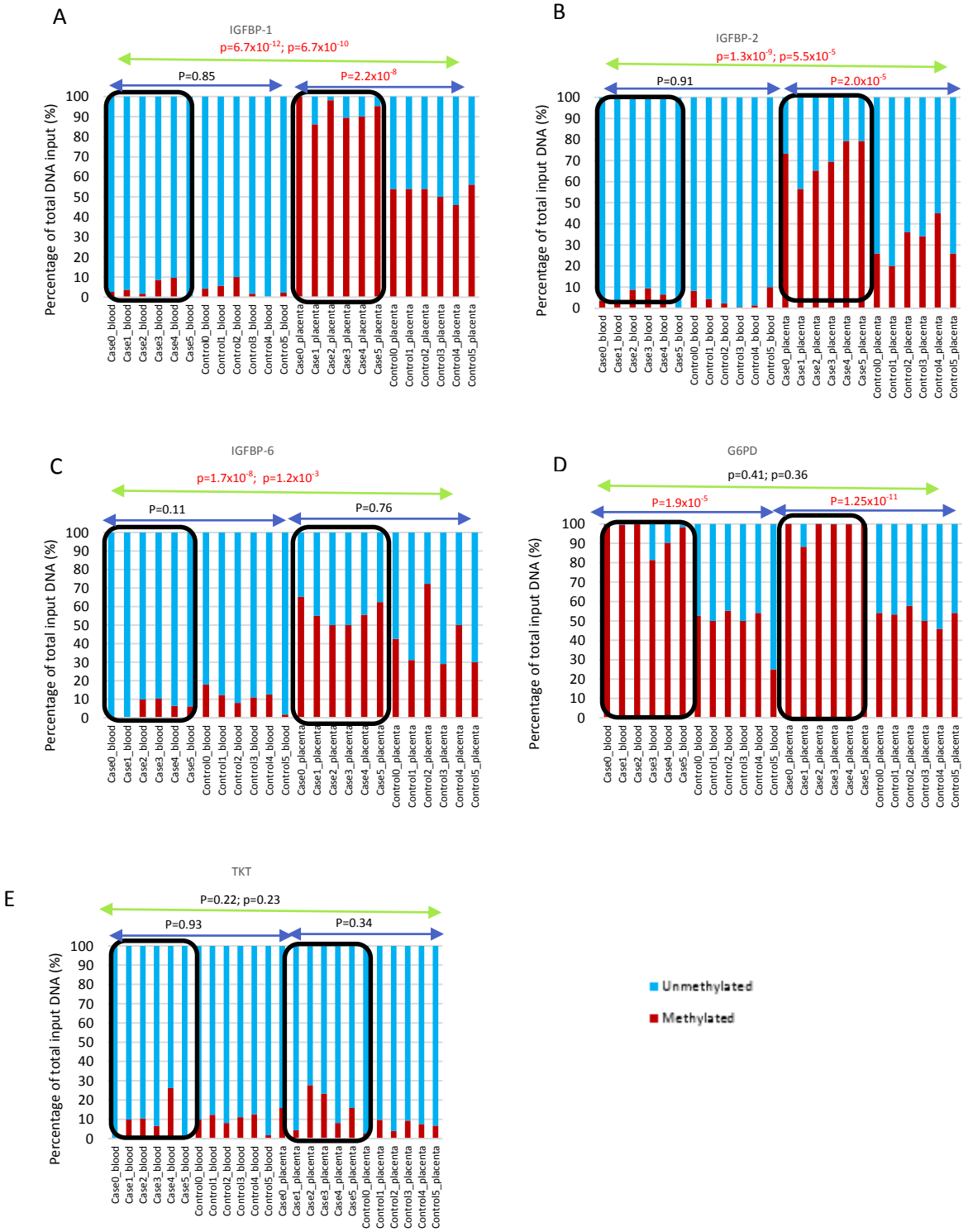


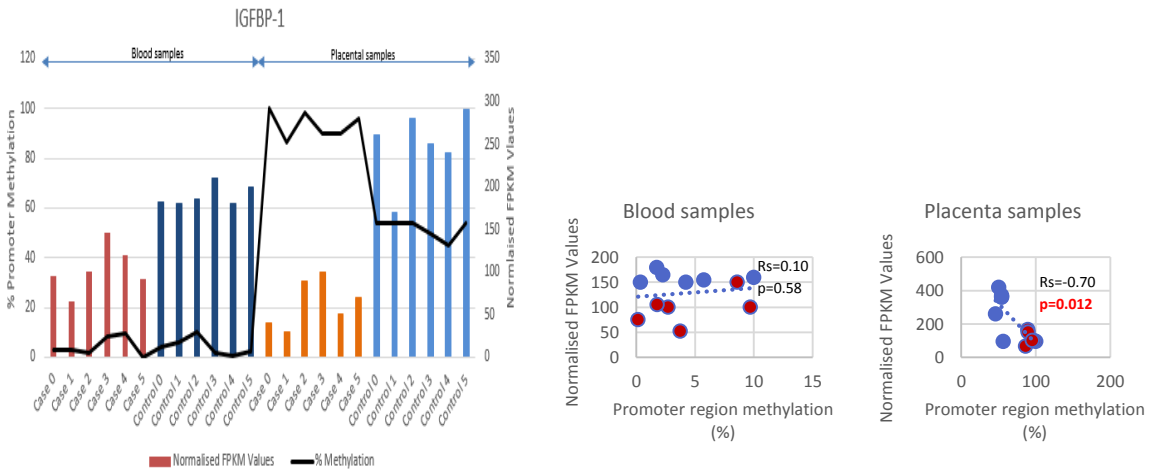
Figure 3.21: Promoter region methylation status of the *IGFBP-1*(A), *IGFBP-2* (B), *IGFBP-6* (C), *G6PD* (D) and *TKT* (E) gene in DNA from maternal case and control blood samples and DNA from exposed and unexposed placental samples. Case blood and placenta samples are enclosed in the black blocks. Blue arrows indicate p-

values for comparing cases and controls and the green arrow indicates p-values for comparing methylation between blood and placenta.

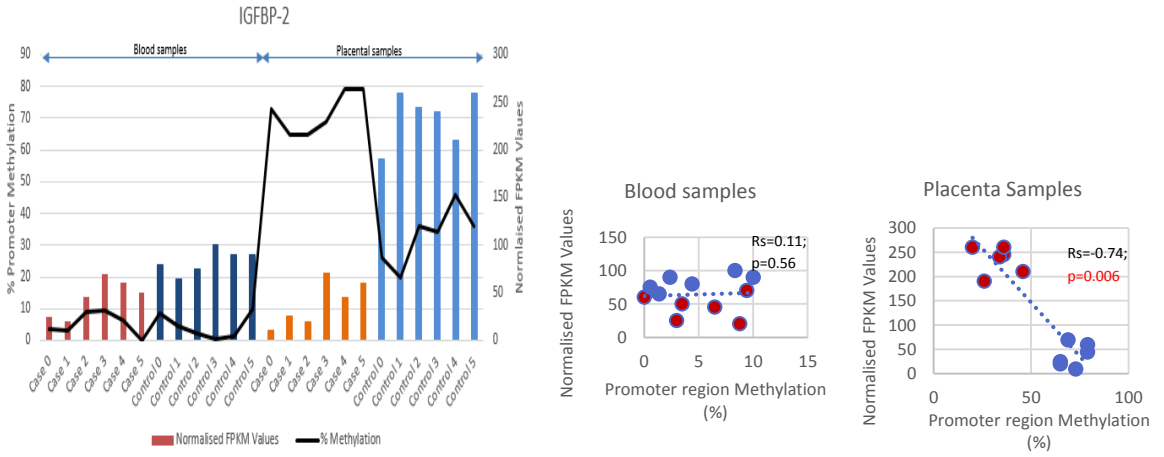
3.7.2 Understanding the relationship between promoter region methylation and mRNA expression of *IGFBP-1*, *IGFBP-2*, *IGFBP-6*, *G6PD* and *TKT*

There is a significant negative association that exists between the level of promoter region methylation and the magnitude of gene expression for *IGFBP-1* (p=0.012) and *IGFBP-2 (*p=0.006*)* in the placental samples (Figure 3.22 A and B). A higher gene expression is associated with lower promoter region methylation. For both genes, this association is absent in the blood samples (mothers)(p=0.58 and p=0.56, respectively). There is no significant association between gene expression and promoter region methylation levels for *IGFBP-6* in the blood or placenta samples (p=0.43 and p=0.65, respectively)(Figure 3.22 C). There is a significant negative association between *G6PD* mRNA expression and promoter region methylation in both the blood (p=0.002) and placenta (p=0.025) samples. A lower expression in the GDM group and exposed placenta is associated with hypermethylation at the promoter region while hypomethylation at the promoter region is associated with increased gene expression (Figure 3.26D). There is no significant association between mRNA expression and promoter region methylation for TKT in both the blood (p=0.74) and placenta (p=0.17) samples (Figure 3.22E).

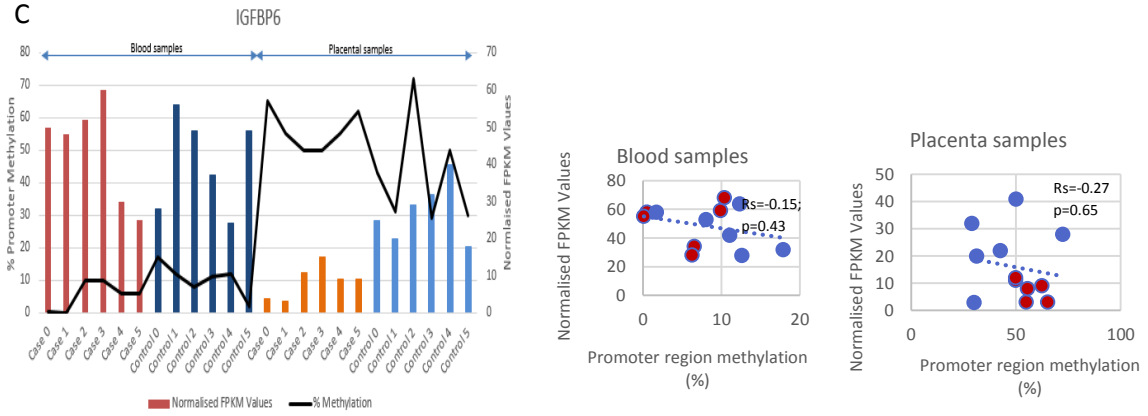
A



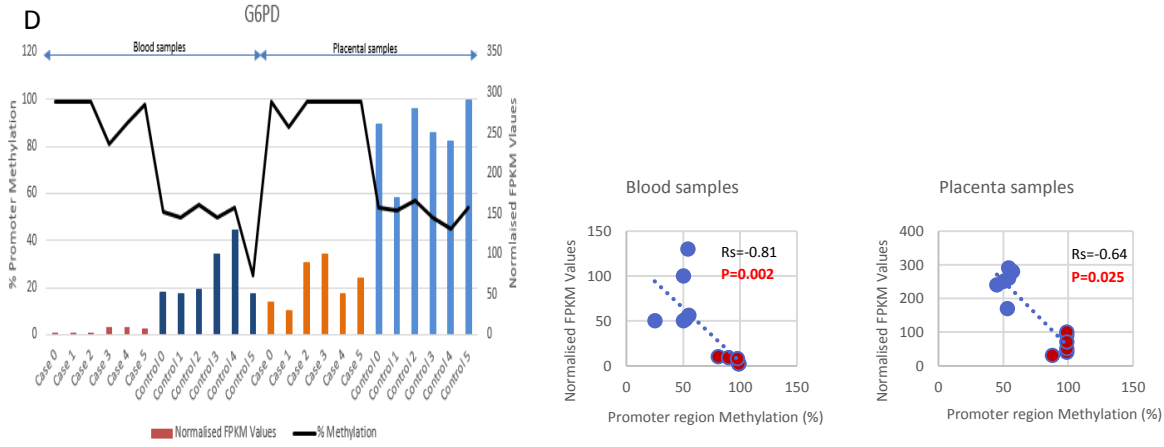
B


IGFBP-2
Blood samples
Placental samples
% Promoter Methylation
Normalised FPKM Vlaues
Normalised FPKM Values
% Methylation
Blood samples
Normalised FPKM Values
Rs=0.11;
p=0.56
Promoter region Methylation (%)
Placenta Samples
Rs=-0.74;
p=0.006
Promoter region Methylation (%)

C


IGFBP6
Blood samples
Placental samples
% Promoter Methylation
Normalised FPKM Vlaues
Normalised FPKM Values
% Methylation
Blood samples
Normalised FPKM Values
Rs=-0.15;
p=0.43
Promoter region methylation (%)
Placenta samples
Rs=-0.27
p=0.65
Promoter region methylation (%)

D


G6PD
Blood samples
Placental samples
% Promoter Methylation
Normalised FPKM Vlaues
Normalised FPKM Values
% Methylation
Blood samples
Normalised FPKM Values
Rs=-0.81
P=0.002
Promoter region Methylation (%)
Placenta samples
Rs=-0.64
P=0.025
Promoter region Methylation (%)

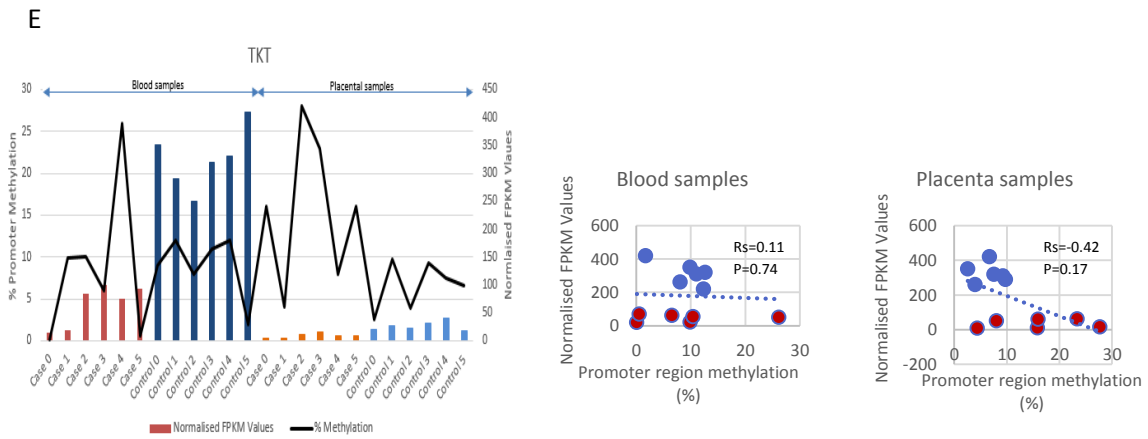


Figure 3.22: **The association between gene expression (bars) and promoter region DNA methylation (black line) in maternal blood and placental samples for *IGFBP-1*(A), *IGFBP-2* (B), *IGFBP-6* (C), *G6PD* (D) and *TKT* (E).** Spearman's correlation rank for each gene IG*FBP-1*(A), *IGFBP-2* (B), *IGFBP-6* (C), *G6PD* (D) and *TKT* (E) is also indicated. Red dots represent case samples and blue dots represent control samples.

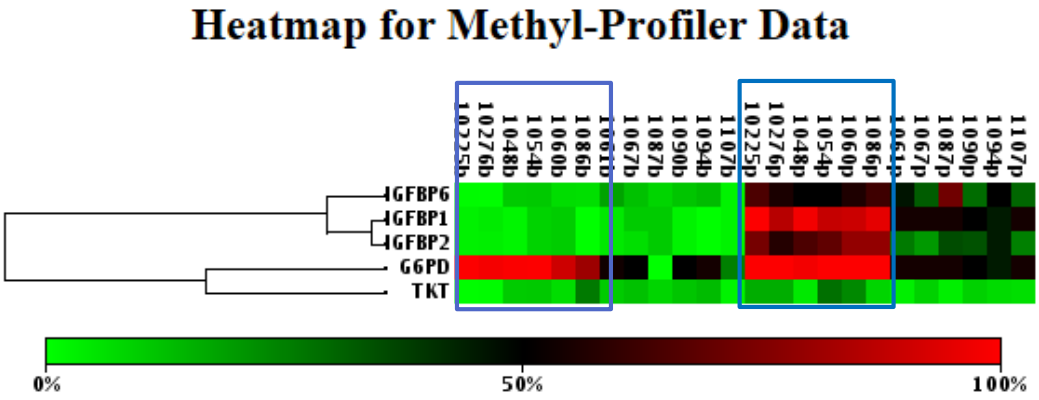


Figure 3.23: **Heat map illustrating the levels of promoter region DNA methylation in selected genes for blood (case and control) and placenta (exposed and unexposed) samples**. Maternal case and exposed placenta samples are enclosed in blue blocks.

The heat map (Figure 3.23) illustrates the higher promoter region methylation of *IGFBP-1*, *IGFBP-2* and *IGFBP-6* in the placenta exposed to GDM *in utero* environment in comparison to the control samples. This figure also highlights the higher level of *G6PD* promoter region methylation in the women who develop GDM in comparison to the control samples.

3.8 Performing Spearman's rank correlation analysis between the mRNA expression values obtained for *IGFBP-1*, *IGFBP-2*, *IGFBP-6*, *G6PD* and *TKT* with other variables

3.8.1 Correlation determination between maternal glucose levels and candidate gene mRNA expression levels in maternal blood

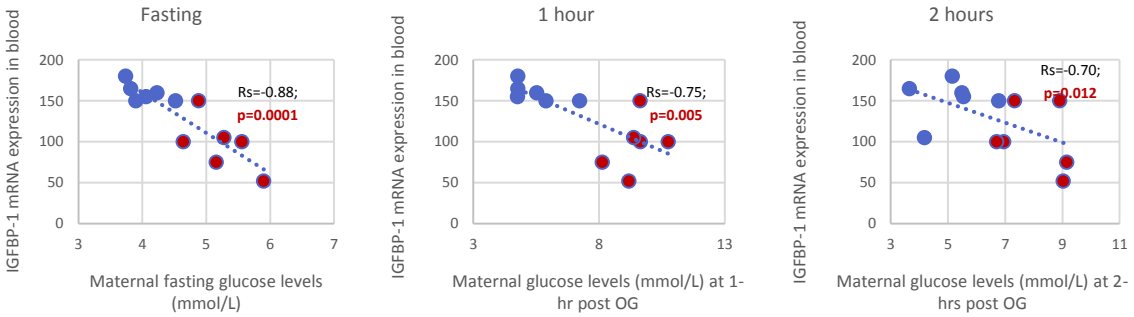
Table 3.23: **Maternal blood glucose levels (mmol/L) at fasting; 1-hr and 2-hrs post oral glucose (OG) intake and the corresponding levels of mRNA expression for *IGFBP-1*, *IGFBP-2*, *IGFBP-6*, *G6PD* and *TKT* in maternal blood.**

	Maternal Glucose levels			**Normalised FPKM Values**				
Sample	Fasting	1-hr OG*	2-hr OG*	**IGFBP-1**	**IGFBP-2**	**IGFBP-6**	**G6PD**	**TKT**
Case0_1054	5.56	9.65	6.95	100	25	58	2	20
Case1_1048	5.90	9.19	9.03	52	20	55	2	22
Case2_1060	5.28	9.38	4.17	105	45	59	3	53
Case3_1086	4.52	9.64	8.91	150	70	68	10	61
Case4_10225	4.64	10.75	6.71	100	60	34	9	50
Case5_10276	5.16	8.14	9.15	75	50	28	8	69
Control0_1061	3.90	7.22	6.77	150	80	32	52	350
Control1_1067	4.06	4.76	5.54	155	65	64	50	220
Control2_1090	4.23	5.52	5.48	160	75	53	56	262
Control3_1094	3.74	4.77	5.14	180	100	42	100	310
Control4_1107	4.88	5.89	7.33	150	90	28	130	320
Control5_1087	3.82	4.77	3.64	165	90	58	50	420

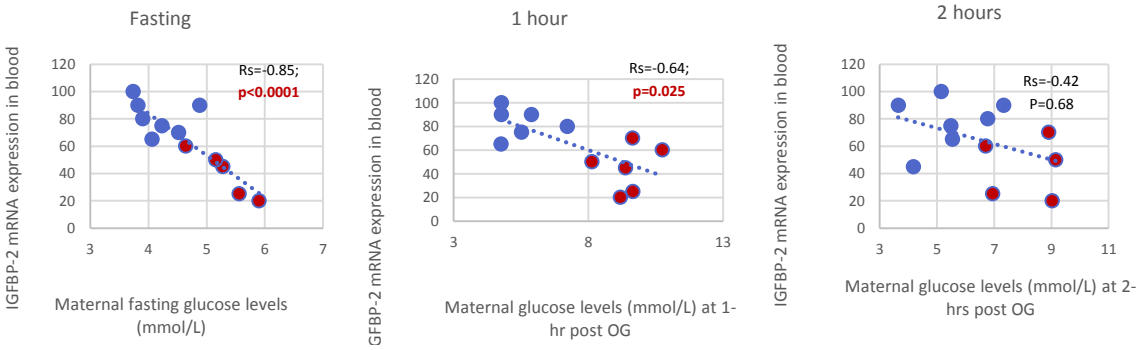
* Indicates the maternal glucose levels measured 1-hr and 2-hrs after the ingestion of 75g anhydrous glucose. OG: Oral glucose

Table 3.23 lists the normalised FPKM values, which represent mRNA expression levels, for *IGFBP-1*, *IGFBP-2*, *IGFBP-6*, *G6PD* and *TKT* in maternal blood and the maternal glucose levels (at fasting; 1-hr and 2-hr post oral glucose (OG)). These data were used to perform Spearman's rank correlation analysis between the variables. For the IGFBPs, significant negative correlations exist between the mRNA expression levels of *IGFBP-1* in maternal blood and maternal glucose levels at fasting (p=0.0001); at 1-hr (p=0.005); and 2-hrs (p=0.012) post OG (Figure 3.24 A); and for *IGFBP-2* at fasting (p=0.0001); and 1-hr post OG (p=0.025) (Figure 3.24 B). No significant correlation was observed between mRNA expression levels and maternal glucose levels 2-hr post OG for *IGFBP-2.* No significant correlation was observed between maternal glucose levels and mRNA expression levels for *IGFBP-6* (Figure 3.24 C). For both *G6PD* and *TKT*, significant negative correlations were observed between the level of mRNA expression and maternal glucose levels at fasting (p=0.006 and p= 0.016, respectively) and at 1-hr post OG (p=0.001 and p=0.005, respectively)(Figure 3.24D and E).

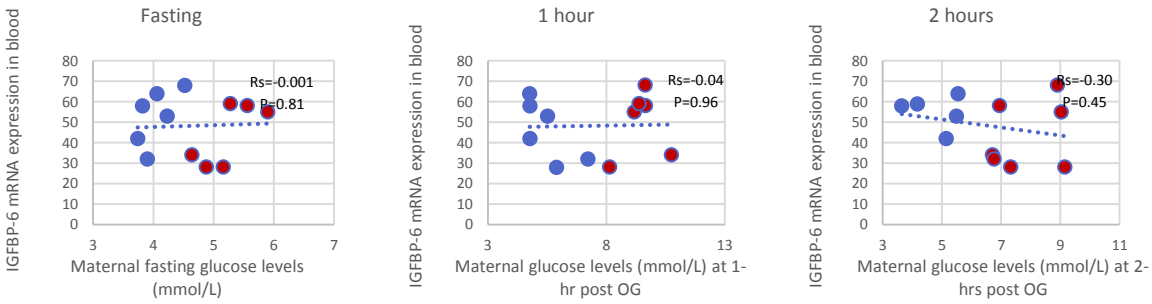
A


Fasting
IGFBP-1 mRNA expression in blood
Rs=-0.88;
p=0.0001
Maternal fasting glucose levels (mmol/L)
1 hour
IGFBP-1 mRNA expression in blood
Rs=-0.75;
p=0.005
Maternal glucose levels (mmol/L) at 1-hr post OG
2 hours
IGFBP-1 mRNA expression in blood
Rs=-0.70;
p=0.012
Maternal glucose levels (mmol/L) at 2-hrs post OG

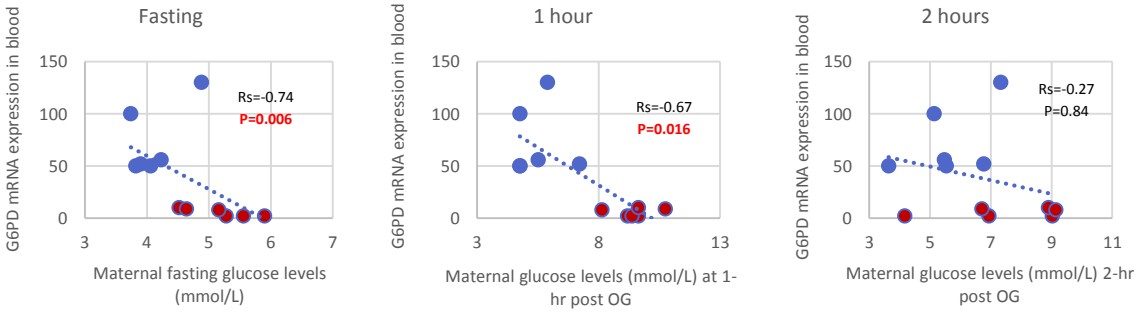
B


Fasting
IGFBP-2 mRNA expression in blood
Rs=-0.85;
p<0.0001
Maternal fasting glucose levels (mmol/L)
1 hour
GFBP-2 mRNA expression in blood
Rs=-0.64;
p=0.025
Maternal glucose levels (mmol/L) at 1-hr post OG
2 hours
IGFBP-2 mRNA expression in blood
Rs=-0.42
P=0.68
Maternal glucose levels (mmol/L) at 2-hrs post OG

C


Fasting
IGFBP-6 mRNA expression in blood
Rs=-0.001
P=0.81
Maternal fasting glucose levels (mmol/L)
1 hour
IGFBP-6 mRNA expression in blood
Rs=-0.04
P=0.96
Maternal glucose levels (mmol/L) at 1-hr post OG
2 hours
IGFBP-6 mRNA expression in blood
Rs=-0.30
P=0.45
Maternal glucose levels (mmol/L) at 2-hrs post OG

D


Fasting
G6PD mRNA expression in blood
Rs=-0.74
P=0.006
Maternal fasting glucose levels (mmol/L)
1 hour
G6PD mRNA expression in blood
Rs=-0.67
P=0.016
Maternal glucose levels (mmol/L) at 1-hr post OG
2 hours
G6PD mRNA expression in blood
Rs=-0.27
P=0.84
Maternal glucose levels (mmol/L) 2-hr post OG

E

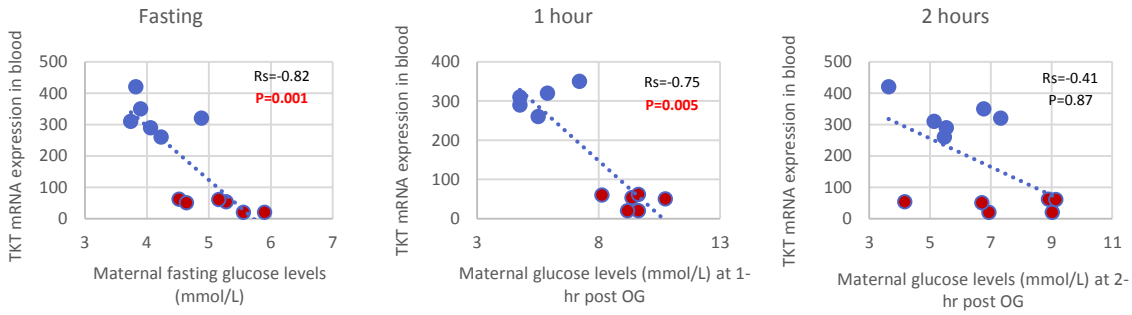


Figure 3.24: **Spearman's rank correlation between maternal glucose levels (mmol/L) and normalized mRNA expression of *IGFBP-1, IGFBP-2, IGFBP-6, G6PD and TKT* in maternal blood**. The red dots indicate GDM cases and the blue dots indicate control samples.

3.8.2 Correlation determination between maternal glucose levels and candidate gene mRNA expression levels in the placenta

Table 3.24: **Maternal blood glucose levels at fasting, 1-hr and 2-hrs post OG (Oral Glucose)(mmol/L) and the corresponding levels of mRNA expression for *IGFBP-1*, *IGFBP-2*, *IGFBP-6*, *G6PD* and *TKT* in the placenta**

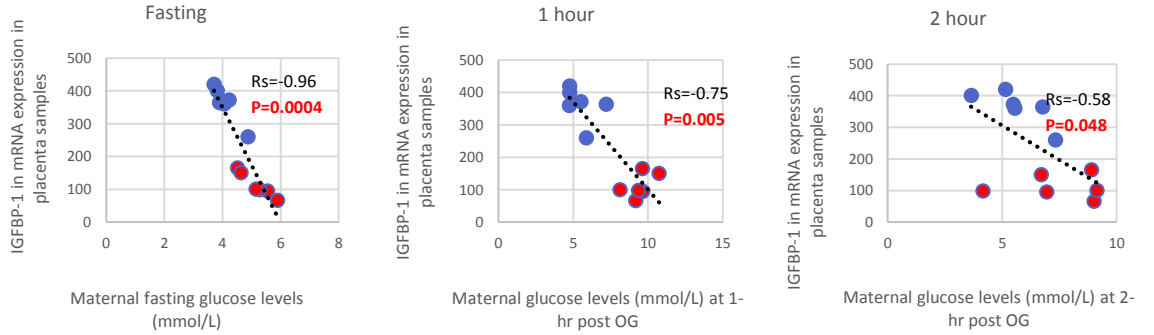
	Maternal Glucose levels			Normalised FPKM Values in the placenta				
Sample	Fasting	1-hr OG*	2-hr OG*	IGFBP-1	IGFBP-2	IGFBP-6	G6PD	TKT
Case0_1054	5.56	9.65	6.95	95	10	3	40	10
Case1_1048	5.90	9.19	9.03	66	25	3	30	9
Case2_1060	5.28	9.38	4.17	98	20	11	90	15
Case3_1086	4.52	9.64	8.91	165	70	12	100	61
Case4_10225	4.64	10.75	6.71	150	45	8	50	50
Case5_10276	5.16	8.14	9.15	100	60	9	70	60
Control0_1061	3.90	7.22	6.77	364	190	22	260	350
Control1_1067	4.06	4.76	5.54	360	260	20	170	290
Control2_1090	4.23	5.52	5.48	372	245	28	280	260
Control3_1094	3.74	4.77	5.14	420	240	32	250	310
Control4_1107	4.88	5.89	7.33	260	210	41	240	320
Control5_1087	3.82	4.77	3.64	95	10	3	290	420

* Indicates the maternal glucose levels measured 1-hr and 2-hrs after the ingestion of 75g anhydrous glucose. OG: Oral glucose

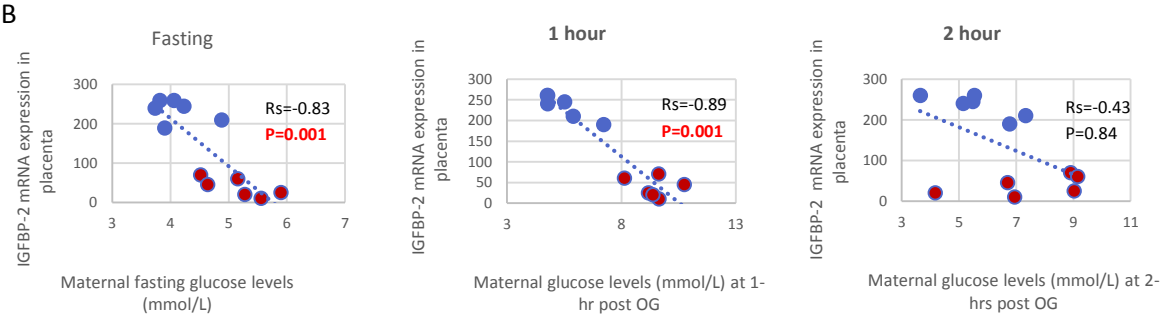
For the placenta, significant negative correlations exist between the mRNA expression levels of *IGFBP-1* and maternal glucose levels at fasting (p=0.0004); at 1-hr (p=0.005); and 2-hrs (p=0.05) post OG (Figure 3.25 A); for *IGFBP-2* at fasting (p=0.001 and 1-hr post OG (p=0.001) (Figure 3.25 B) and for *IGFBP-6* at fasting (p=0.008) and at 1-hr (p=0.012) post OG (Figure 3.25

C). For *G6PD* significant negative correlations were observed between the level of mRNA expression and maternal glucose levels at fasting (p=0.001); at 1-hr post OG (p=0.007); and at 2-hr post OG (p=0.04)(Figure 3.25 D). No significant correlation was observed between maternal glucose levels and mRNA expression levels for *TKT* in the placenta (Figure 3.25 E).

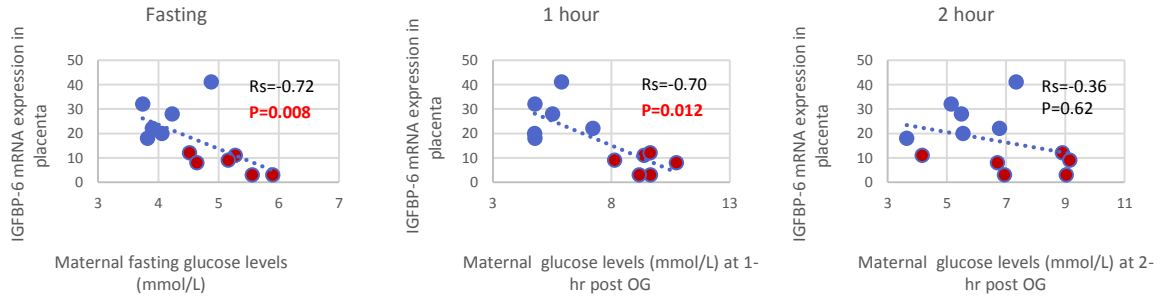
A



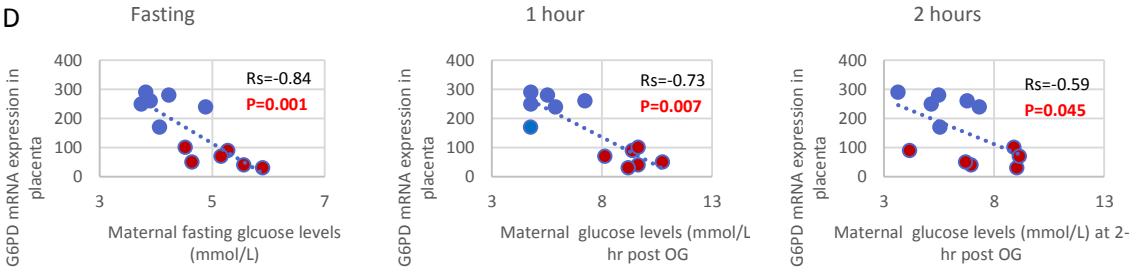
B



C



D



E

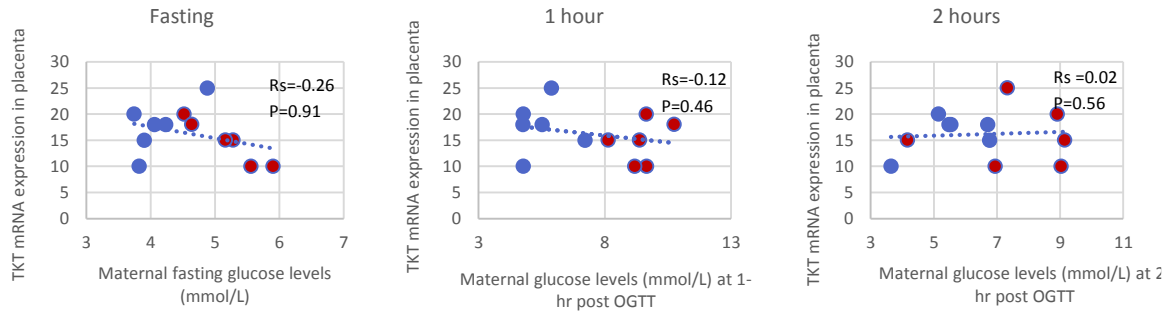


Figure 3.25: **Spearman's rank correlation between maternal glucose levels (mmol/L) and normalized mRNA expression of *IGFBP-1, IGFBP-2, IGFBP-6, G6PD and TKT* in the placenta.** The red dots indicate GDM cases and the blue dots indicate control samples.

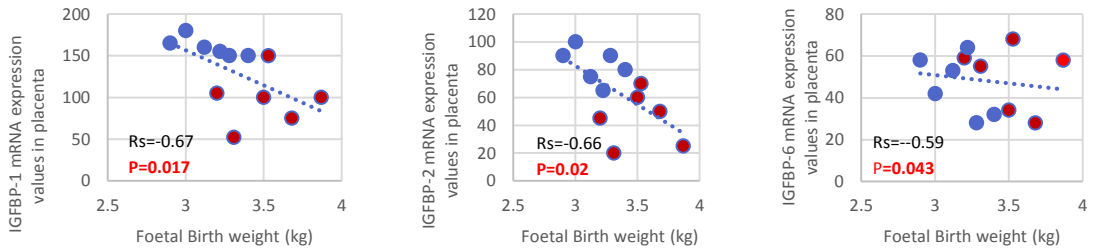
3.8.3 Correlation determination between mRNA expression of the IGFBPs in maternal blood and placental samples with foetal birth weight and maternal BMI

IGFBPs are known to play a role in foetal growth (Agrogiannis et al., 2014, Duron et al., 2012, Gupta, 2015, Gupta et al., 2015) and for this reason, we wanted to observe whether there was an association between *IGFBP-1*, *IGFBP-2* and/or *IGFBP-6* mRNA expression levels and foetal birth weight in this study. Table 3.25 lists the BMI (kg/m^2) of the mother as well as the birth weight of the corresponding foetus. The mRNA expression levels of the IGFBP genes in the blood and placenta are listed in Table 3.23 and Table 3.24 respectively. These data were used to perform Spearman's rank correlation analysis between the variables.

Table 3.25: Maternal BMI and the birth weight (kg) of her neonate

Sample	Birth weight (kg)	BMI
Case0_1054	3.87	35.4
Case1_1048	3.31	36.5
Case2_1060	3.20	36.1
Case3_1086	3.53	36.2
Case4_10225	3.50	48.5
Case5_10276	3.68	35.0
Control0_1061	3.40	24.0
Control1_1067	3.22	23.4
Control2_1090	3.12	35.3
Control3_1094	3.01	29.7
Control4_1107	3.28	38.6
Control5_1087	2.89	34.0

A



B

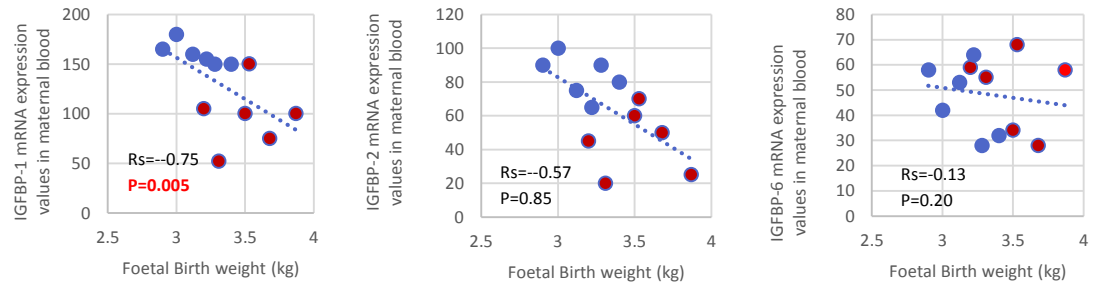


Figure 3.26: **Spearman's rank correlation (Rs) between foetal birth weight (kg) and normalized I*GFBP-1, IGFBP-2 and IGFBP-6* mRNA expression (measured via RNA-seq) in the placenta (A) and in maternal blood (B).** The red dots indicate case samples and blue dots indicate control samples.

There was a significant correlation between birth weight and *IGFBP-1* (p=0.017); *IGFBP-2* (p=0.02) and *IGFBP-6* (p=0.04) mRNA expression levels in the placenta. A lower expression of these genes is associated with larger birth weight (Figure 3.26 A). There was also a significant negative correlation between *IGFBP-1* mRNA expression in maternal blood and foetal birth weight (p=0.005). There was no significant correlation observed between *IGFBP-2* and *IGFBP-6* in maternal blood and foetal birth weight (Figure 3.26 B). Although there appears to be a trend towards a lower expression of IGFBPs in women with higher BMI, this association is not significant in our cohort for any of the IGFBPs studied (Figure 3.27).

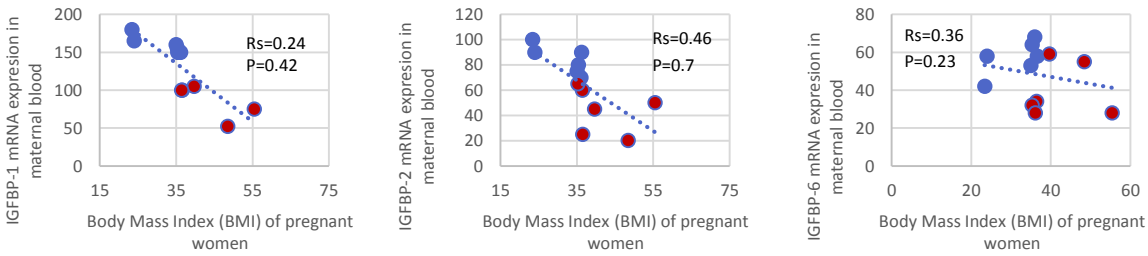


Figure 3.27: **Spearman's rank correlation (Rs) between normalized *IGFBP-1, IFGBP-2* and *IGFBP-6* mRNA expression in maternal blood and Body Mass Index (BMI).** The red dots indicate case samples and blue dots indicate control samples.

3.8.4 Correlation between maternal glucose levels and DNA methylation in maternal blood and placental tissues

There was no significant correlation observed between promoter region methylation of *IGFBP-1*, *IGFBP-2* and *IGFBP-6* in maternal blood and maternal glucose levels (Figure 3.28 A, 3.29 A and 3.30 A, respectively). However, in the placenta (Figure 3.28 B), there appears to be a significant positive correlation between *IGFBP-1* promoter region methylation and maternal glucose levels at fasting (p=0.03) and 1-hr post OG (p=0.005) but not at 2-hrs. For *IGFBP-2*, a significant positive correlation was observed between promoter region methylation in the placenta and maternal glucose levels at fasting and 1-hr post OG (p=0.01 and p=0.001 respectively, Figure 3.29 B) but not at 2-hrs. For *IGFBP-6*, a significant positive correlation was observed between promoter region methylation in the placenta and maternal glucose levels at fasting (p=0.02) and 1-hr post OG (p=0.05)(Figure 3.30B). For *G6PD,* a significant positive correlation was observed in maternal blood (Figure 3.31 A) and placental tissue (Figure 3.31 B) at fasting (p=0.001 and p=0.02, respectively) and 1-hr post OG (p=0.001 and p=0.004, respectively). This association was not observed at 2-hr post OG in either the blood or placenta tissue (Figure 3.31 A and B, respectively). No correlation between promoter region methylation and maternal glucose levels were observed for *TKT* in either maternal blood or placenta samples (Figure 3.32 A and B, respectively).

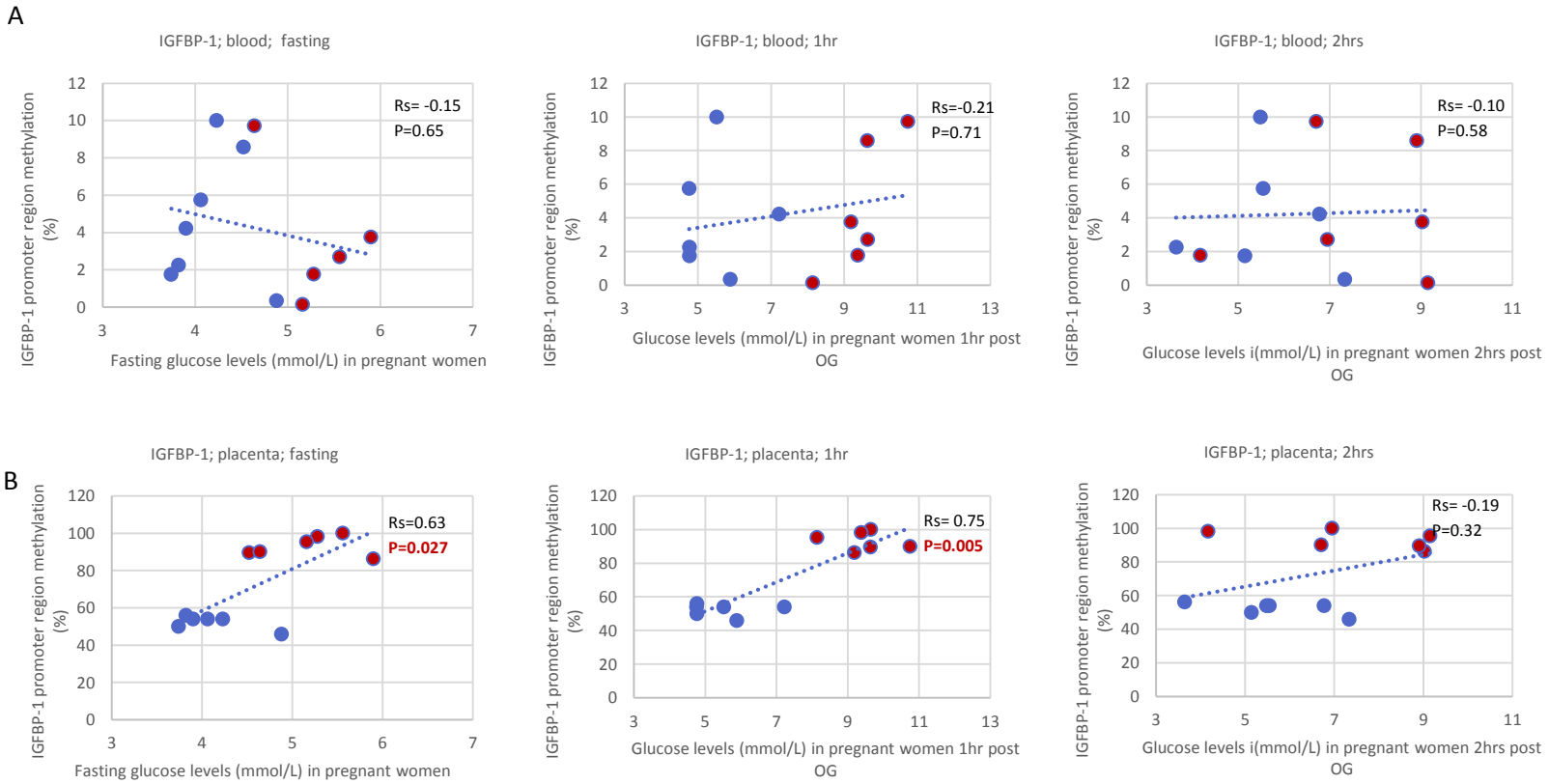


Figure 3.28. Spearman's rank correlation between blood (A) and placental (B) *IGFBP*-1 gene promoter DNA methylation and maternal glucose levels (fasting; 1-hr post OG and 2-hrs post OG). Red dots indicate case samples and blue dots indicate control samples.

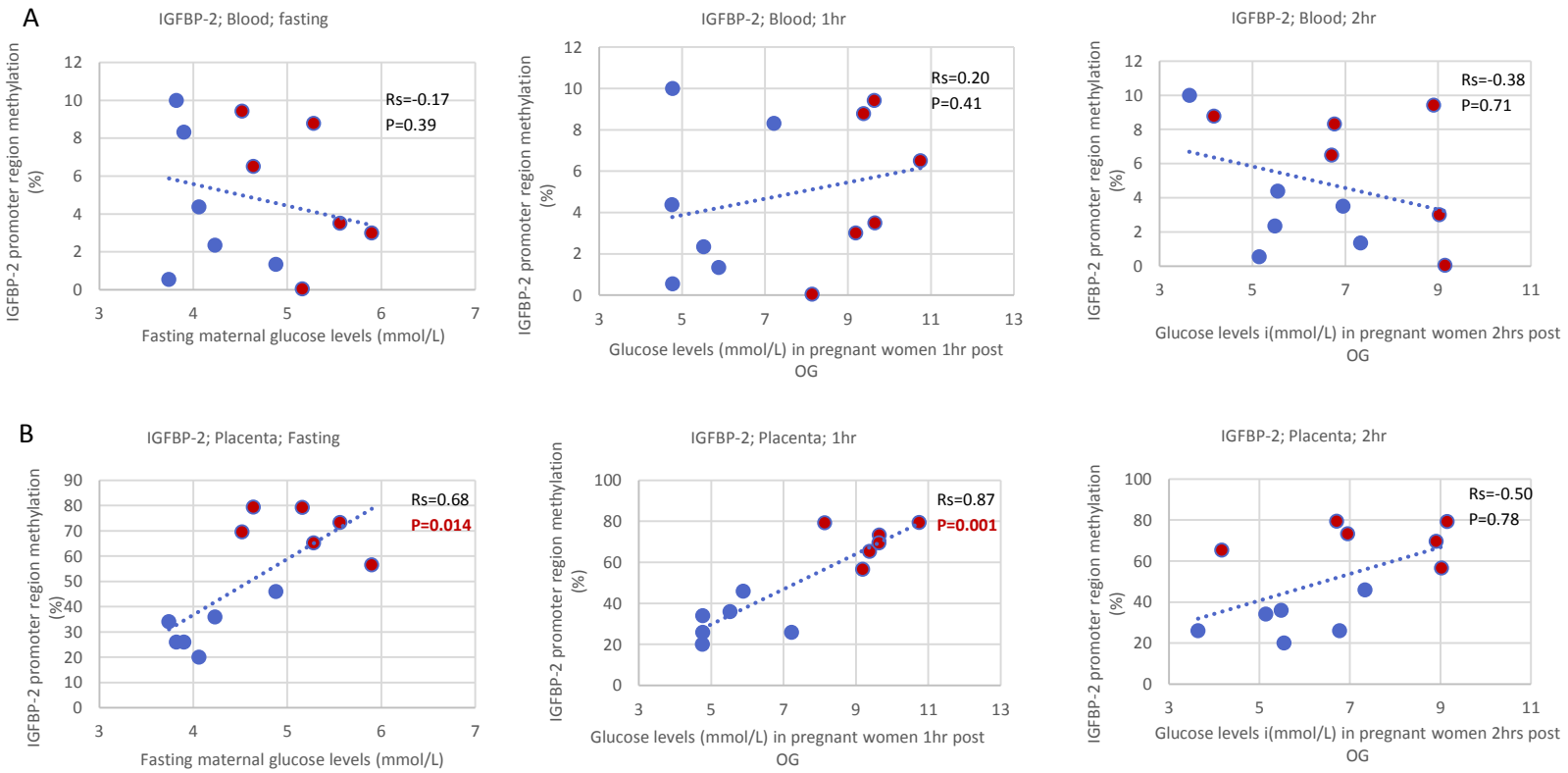


Figure 3.29. Spearman's rank correlation between maternal blood (A) and placental (B) *IGFBP-2* gene promoter DNA methylation and maternal glucose levels (fasting; 1-hr post OG and 2-hr post OG). Red dots indicate case samples and blue dots indicate control samples.

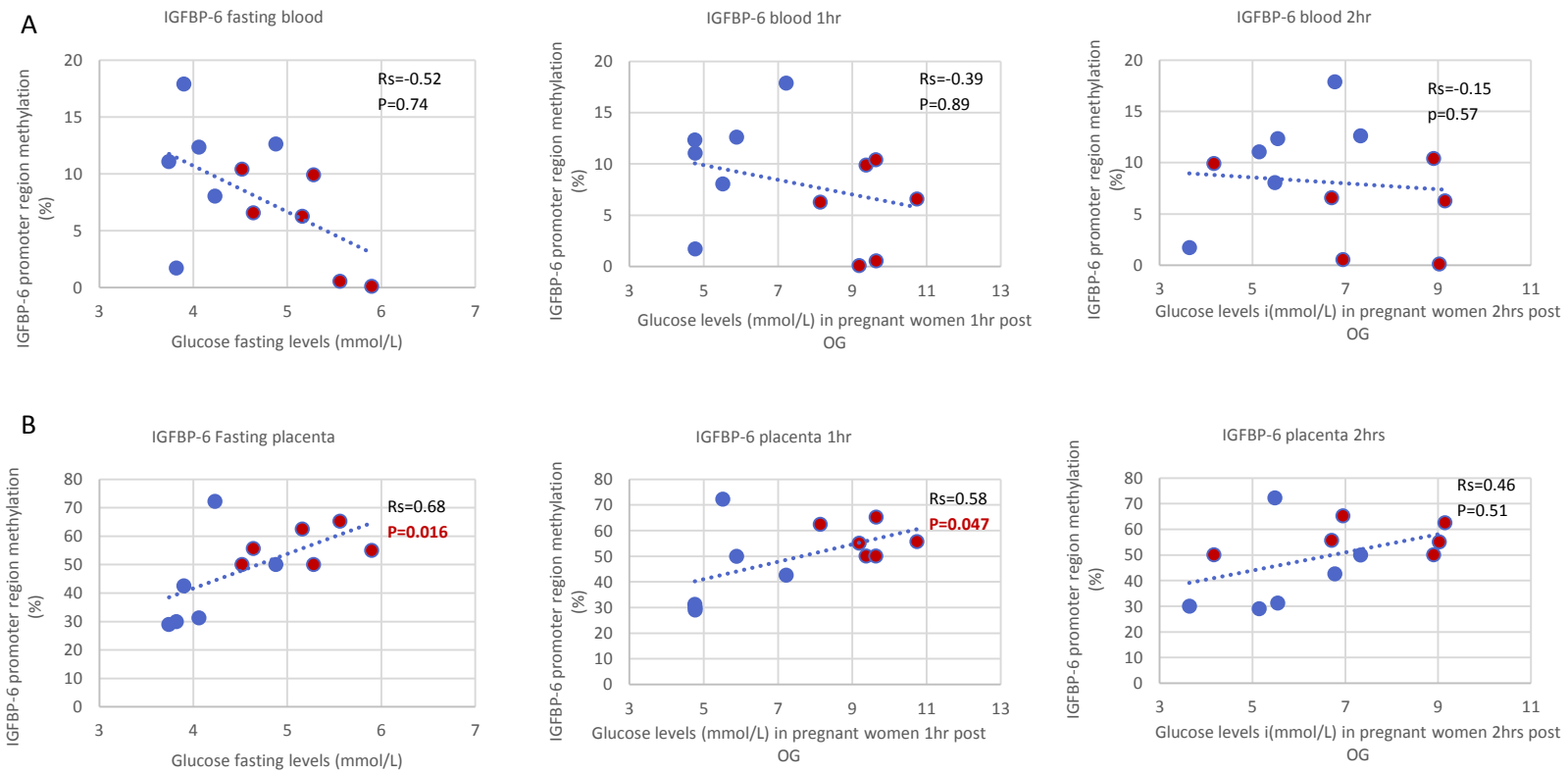


Figure 3.30. Spearman's rank correlation between maternal blood (A) and placental (B) *IGFBP-6* gene promoter DNA methylation and maternal glucose levels (fasting; 1-hr post OG and 2-hr post OG). Red dots indicate case samples and blue dots indicate control samples.

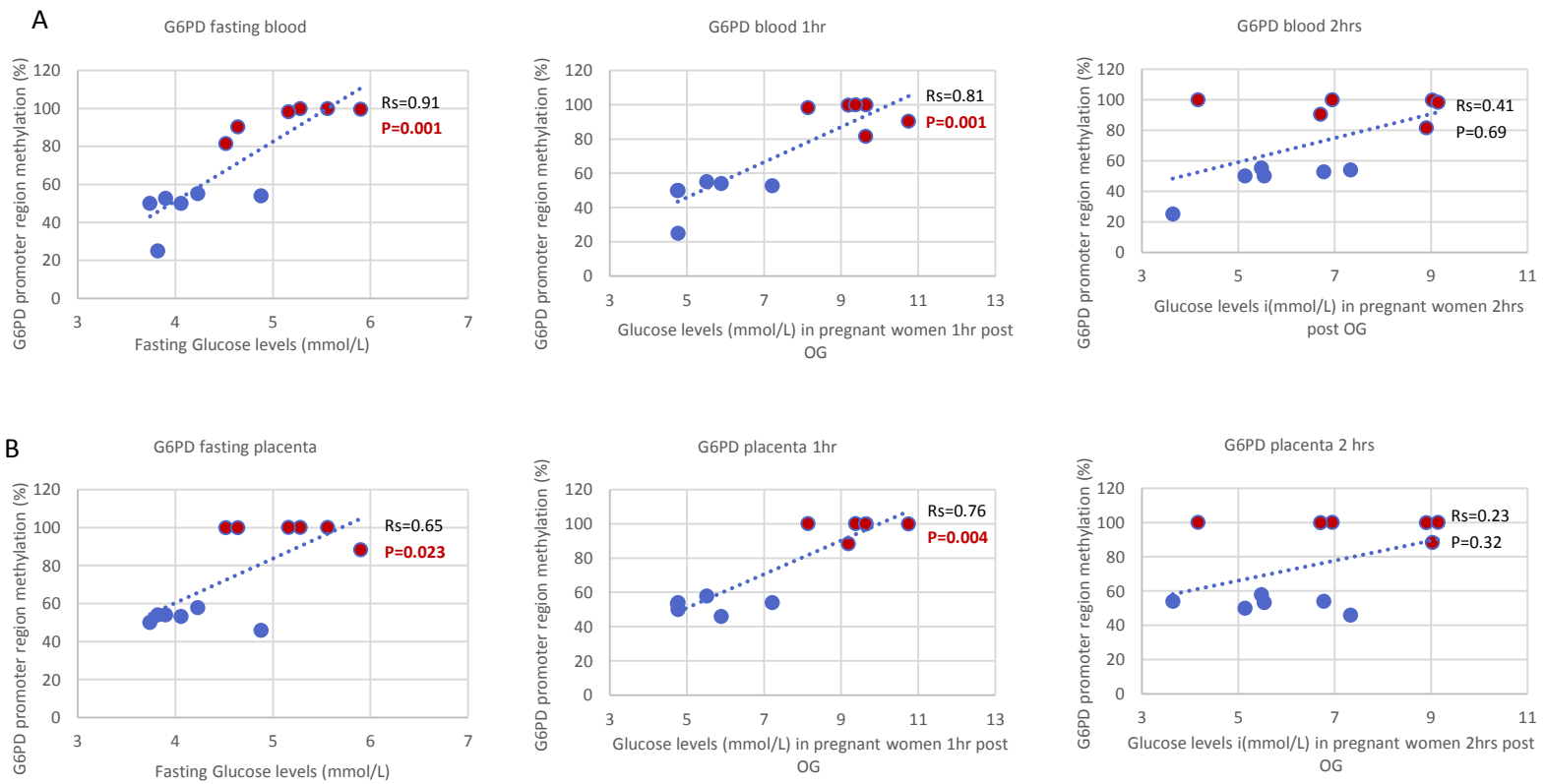


Figure 3.31 Spearman's rank correlation between maternal blood (A) and placental (B) *G6PD* gene promoter DNA methylation and maternal glucose levels (fasting; 1-hr post OG and 2-hr post OG). Red dots indicate case samples and blue dots indicate control samples.

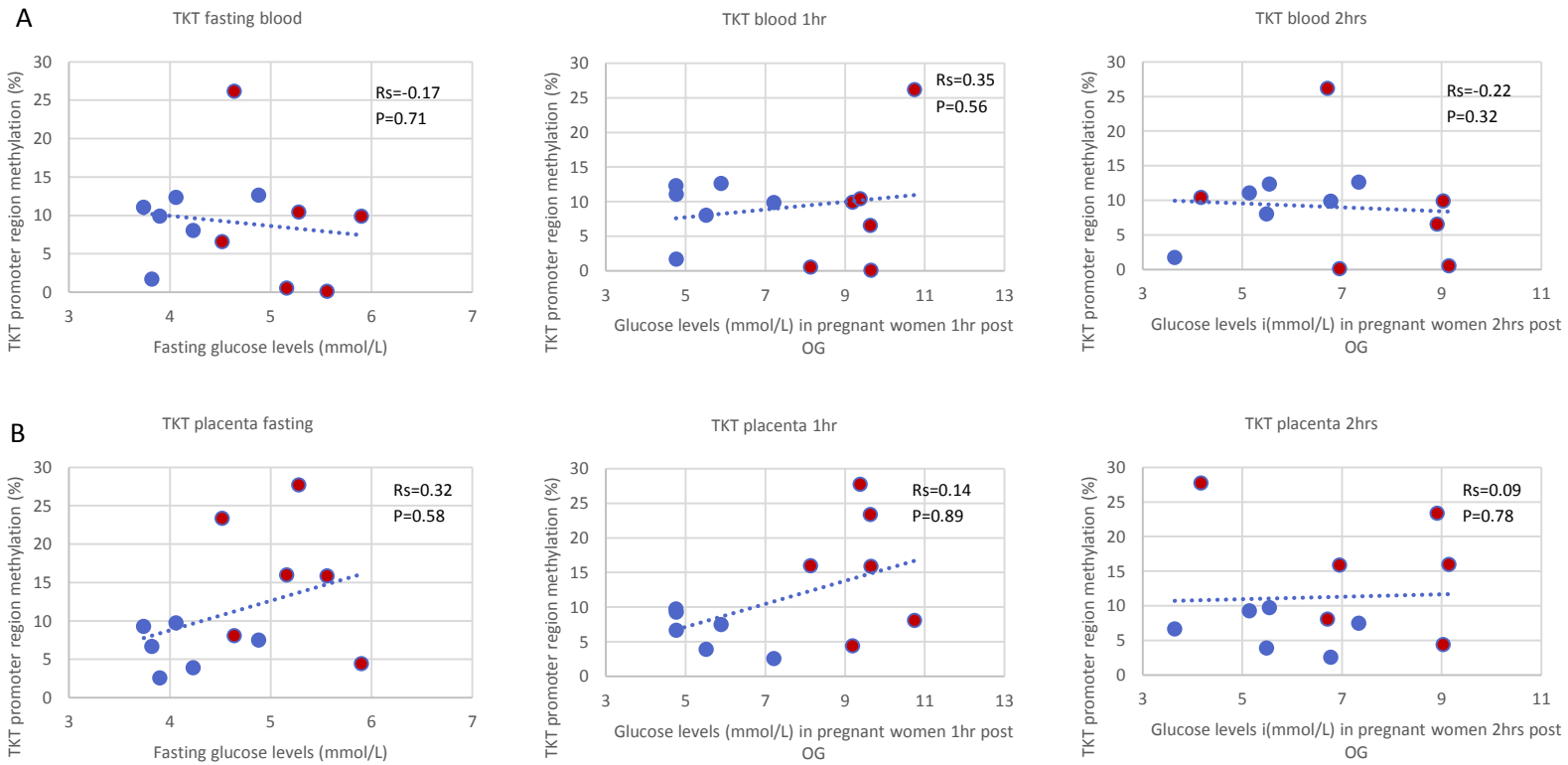


Figure 3.32. Spearman's rank correlation between maternal blood (A) and placental (B) *TKT* gene promoter DNA methylation and maternal glucose levels (fasting; 1-hr post OG and 2-hr post OG). Red dots indicate case samples and blue dots indicate control samples.

3.8.5 Correlation between foetal birth weight and IGFBP promoter region methylation in the placenta and maternal blood samples

There is a significant positive correlation between foetal birth weight and the level of IGFBPs (1, 2 and 6) promoter region methylation in the placenta. The higher the level of methylation at the promoter region, the higher the birth weight (Figure 3.33 A). There was no significant correlation observed between *IGFBP-1*, *IGFBP-2* and *IGFBP-6* promoter region methylation in the mothers blood and birth weight (Figure 3.33 B).

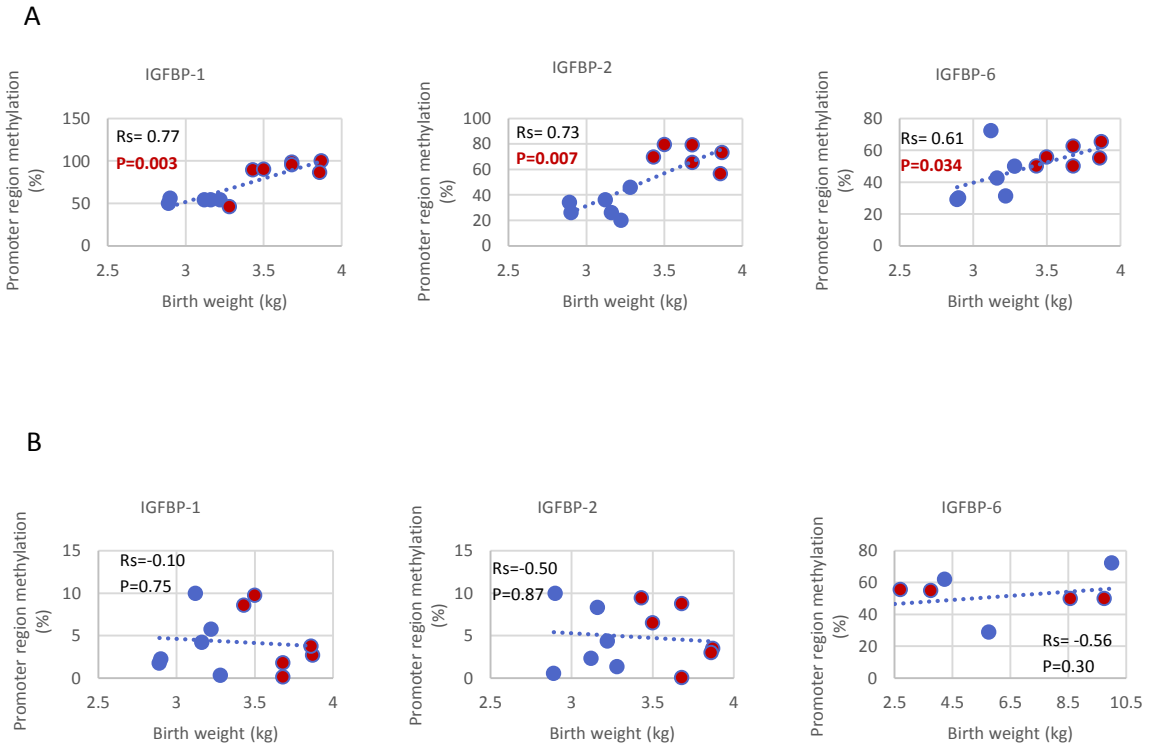


Figure 3.33. Spearman's rank correlation between methylation levels in the promoter region of IGFBP-1, 2 and 6 in placental (A) and maternal blood (B) samples and birth weight. Red dots indicate case samples and blue dots indicate control samples.

CHAPTER 4

4. DISCUSSION

Gestational diabetes mellitus, an important health issue that is increasing in prevalence every year (Dabalea et al., 2005), creates an adverse *in utero* environment for the developing foetus and increases the risk of developing T2D in the foetus as well as the mother. Therefore, GDM is a good model to study the mechanisms involved in foetal metabolic programming and also to elucidate new mechanisms to help diagnose, treat and prevent its consequences for the offspring (as newborns and later as adults) as well as for successive generations. Currently there is little molecular understanding of how foetal programming may occur (Finer et al., 2015). Epigenetic modification is currently a very promising mechanism to explain foetal metabolic programming and many studies are beginning to shed light on the mechanisms for the intergenerational transmission of disease risk (Bouchard et al., 2012). Epigenetic control of gene expression is a key step in understanding the development of a particular phenotype. Biological conditions and disease status are known to be largely characterized by differences in gene expression levels. Epigenetic mechanisms provide a component of plasticity that allows for adaptation during times of early environmental stresses such as prenatal over-nutrition and under-nutrition (Joss-Moore and Lane, 2012).

Gene methylation in placental tissue is generally lower when compared to other somatic tissues (Christensen et al., 2009, Nawathe and Lees, 2016, Nawathe et al., 2016). This lower level of methylation has been associated with promoting healthy foetal development throughout gestation. Placental function and the intrauterine environment play critical roles in foetal programming (Nawathe and Lees, 2016). Altered DNA methylation in the placenta plays a significant role in optimal placental and foetal growth (Serman et al., 2007). Candidate gene studies have demonstrated an association between DNA methylation in placental tissue and maternal hyperglycemia (Bouchard et al., 2010, Nawathe and Lees, 2016, Nawathe et al., 2016). Bouchard et al. (2010) observed an association between maternal hyperglycemia with lower DNA methylation in placental tissue at the *LEP* gene (Bouchard et al., 2010) as well as higher levels of methylation at the *ADIPOQ* gene in foetal placental tissue (Bouchard et al.,

2012). Associations have also been observed between maternal hyperglycemia and lower DNA methylation at the promoter region of other candidate genes involved in metabolism such as *ABCA1* (in cord blood), *LPL* (in placenta) and in *IGFBP3* (maternal and cord blood) (Houde et al., 2013, Houde et al., 2014).

In pregnancy, many adaptions occur to ensure a healthy metabolic balance between the mother and foetus while ensuring proper foetal development. In the context of glucose metabolism, these adaptions occur to ensure that sufficient amounts of glucose reach the foetus to promote normal development. In GDM, although insulin sensitivity is only slightly decreased compared with pregnant women with normal glucose tolerance, insulin secretion in women with GDM is significantly decreased (Angueira et al., 2015). This results in the accumulation of glucose in the mother's blood. In normal pregnancies, the ß-cells of the pancreas will up-regulate insulin secretion in response to the increasing concentration of glucose, but in a certain percentage of women, this up-regulation does not take place. This results in GDM.

In this study I used RNA sequencing to study gene expression as it produces a large dataset that is useful both in identifying genes that are differentially expressed between cases and controls in the blood of the mother and in the placenta, and provides a good quantitative estimate of gene expression. This study is not hypothesis driven as is the case in candidate gene studies, but is exploratory, thereby covering the entire range of RNA transcription. Twenty genes with large differential expression were selected as good candidates based on the gene ontology (GO) and enrichment analysis results as well as an extensive literature research. These candidate genes were validated using quantitative real-time PCR and exhibited significant differential expression in the quantitative real-time PCR assays ($p<0.05$), confirming the RNA sequencing results. It is noteworthy to mention that the direction of change (increase/decrease) in expression was 100 % consistent between RNA sequencing and qPCR data for the validated genes. Each case and control participant in the study was matched as closely as possible with regard to ethnicity, age and BMI, variables that are known to affect DNA methylation (Zaghlool et al., 2015, Zhang et al., 2011). Reducing the number of variables between case and control groups ensures that the differential variation in gene expression is

most likely due to the presence or absence of gestational diabetes, rather than confounding factors. Only cases with female offspring were analyzed as methylation patterns of a large number of genes, especially in the placenta, have been shown to be foetal sex-specific (Hall et al., 2014, McCarthy et al., 2014).

We observed that 60 genes were significantly differentially expressed in the mother's blood at the time GDM was diagnosed (24-28 weeks gestation), of which three are well characterized genes that have known associations with T2D and therefore also potentially with GDM (*AKT2*, *KCNQ1* and *GSK3A*). Others included an open reading frame of unknown function that has been mentioned previously as a possible marker for GDM (*C8ORF80*); a mitochondrial glutamate carrier 1 gene (*SLC25A22*) as well as five genes that encode enzymes which function in the pentose phosphate pathway (*G6PD*, *TKT*, *ALDOA*, *DCXR* and *PGLS*). The *G6PD* and *TKT* genes were also significantly under-expressed in exposed placenta. Placental tissue taken from the mothers with GDM contained 52 genes that were significantly differentially expressed, three of which were insulin-like growth factor binding proteins (*IGFBP-1*, *IGFBP-2* and *IGFBP-6)*. The *IGFBP-1* and *IGFBP-2* genes were also under-expressed in the blood of women who develop GDM, however this differential expression only reached levels of significance for *IGFBP-1.* The five genes chosen for promoter region methylation analysis were all significantly down-regulated in either the women who developed GDM (*G6PD, TKT* and *IGFBP-1)* or in exposed placenta (*G6PD, IGFBP-1, IGFBP-2* and *IGFBP-6).*

Significant hypermethylation at the promoter region of *G6PD* was observed in both women who developed GDM as well as in exposed placenta. There was no observed variation in the methylation of the promoter region of *TKT* in either group. Significant hypermethylation at the promoter region of *IGFBP-1* and *IGFBP*-2 was observed in exposed placenta, but not in women with GDM. There appeared to be a trend towards increased promoter region methylation of *IGFBP-6* in exposed placenta, however this did not reach levels of significance when compared to unexposed placenta (Table 4.1).

Table 4.1 The five genes chosen for gene specific promoter region methylation

Candidate Genes	Women with GDM		Exposed placenta	
	mRNA expression	Promoter region methylation	mRNA expression	Promoter region methylation
G6PD	↓	↑	↓	↑
TKT	↓	■	↓	■
IGFBP-1	↓	■	↓	↑
IGFBP-2	↓	■	↓	↑
IGFBP-6	↓	■	↓	↓

Red arrow: Significant change, blue arrow: non-significant change; yellow block: no trend or significant trend observed

4.1 Insulin-like growth factor binding proteins (IGFBPs) mRNA expression, gestational diabetes and foetal growth

Normal foetal development is dependent on a balanced interplay between growth suppressors and promoters originating from foetal, placental and maternal compartments. The IGF axis is a complex system composed of a family of interacting ligands (IGF-1 and IGF-2), two receptors (IGF-1R and IGF-2R) and insulin-like growth factor binding proteins (IGFBPs)(Huang et al., 2015, Lu et al., 2012). This system plays an important role in the regulation of somatic growth in an endocrine manner and in the proliferation and differentiation of normal and malignant cells in a paracrine-autocrine manner (Lee et al., 2016). IGF-1 and -2 are predominately produced by adult and foetal liver although the placenta also expresses these peptides (Hiden et al., 2009). The majority of IGFs in blood and tissues are bound to six of the named IGFBPs that have been identified in humans (Gonzalez-Parra et al., 2002). These IGFBPs form complexes with both IGF-1 and IGF-2 with a high affinity preventing them from binding to their receptors (Allen et al., 2003, Hiden et al., 2009). In this way, the bioavailability of IGFs is controlled by IGFBPs. Therefore, as modulators of IGF actions, IGFBPs assume an important role in the process of foetal growth (Giudice et al., 1995, Lu et al., 2012).

The IGFBPs are globally distributed in all tissues and cells, but the majority of circulating levels of these proteins in humans are produced by the liver under the regulation of IGFs and insulin (Rajpathak et al., 2009). Four of the IGFBPs, (*IGFBP-1*, *IGFBP-2*, *IGFBP-3* and *IGFBP-5*) have been implicated in T1D, obesity and insulin resistance in both animal models and human subjects (Lu et al., 2012, Ruan and Lai, 2010). The IGFBPs have also been identified as surrogate markers for metabolic syndrome, cardiovascular disease and cancers (Lu et al., 2012, Pon et al., 2015, Vasylyeva and Ferry, 2007). In terms of glucose metabolism, IGFBPs play an important role in insulin signalling, enhancing peripheral glucose uptake, decreasing hepatic glucose output and modifying lipid metabolism. A number of studies have demonstrated an association between IGFBPs, glucose tolerance and insulin resistance (Aguirre et al., 2016, Heald et al., 2001). Nawathe et al. (2016) found elevated mRNA and protein levels of IGFBP-1, IGFBP-2, IGFBP-3, IGFBP-4 and IGFBP-7 in placenta of small for gestational age (SGA) neonates and decreased expression of IGFBP-1, IGFBP-2, IGFBP-3 and IGFBP-4 in the large for gestational age (LGA) group. These findings suggest that IGFBPs play a role in the "net IGF bioavailability" in pregnancies affected with SGA and LGA (Nawathe et al., 2016).

The IGFBPs also contain functional domains which enable IGF-independent actions (Bach, 2015c, Bach, 2015a, Bach, 2015b, Wheatcroft and Kearney, 2009). This is a result of their transport into the nucleus where they may exert IGF-independent activities by transcriptional activation of genes. The IGF-independent actions of IGFBPs are not as well understood in comparison to their IGF-dependent actions (Forbes et al., 2012).

4.1.1 *IGFBP-1*

IGFBP-1 is the main insulin-like growth factor binding protein of amniotic fluid and a significant binding protein in maternal and foetal serum (Hills et al., 2013, Holmes et al., 2000). This 25 kDa protein is mainly secreted by the liver, however, it is also produced by granulosa cells and decidualised endometrium (Juul, 2003, Khosravi et al., 2007). The primary physiological function of IGFBP-1 appears to be the regulation of the bioavailability of IGF-1 and IGF-2, although it does also have effects that are IGF-independent. *IGFBP-1* is the only IGFBP acutely

affected by glucoregulatory hormones and may play an active role in glucose modulation. According to many studies, insulin is the main regulator of *IGFBP-1* under basal (non-stress) conditions, exerting a suppressive effect (Gibson et al., 1995, Heald et al., 2001, Loukovaara et al., 2005). Therefore, the production of *IGFBP-1* in the liver is dependent on insulin supply in the portal circulation and the concentration of this binding protein in circulation varies significantly depending on whether the individual is in the non-fed or postprandial state (Heald et al., 2006). In a study performed by Bae et al. (2013) they observed that an increase in insulin resulted in the suppression of hepatic *IGFBP-1* gene expression, and that *IGFBP-1* levels rapidly declined after feeding (Bae et al., 2013).

The levels of IGFBP-1 are decreased in conditions associated with insulin resistance (Aguirre et al., 2016, Sandhu et al., 2002). Kabir et al. (2010) showed that the expression of *IGFBP-1* increases in a fasting state (Kabir, 2014, Lappas, 2015) in an effort to recruit and increase the expression and activity of IGFs. *In vivo* studies have shown that IGFBP-1 injection into rats resulted in a significant increase in glucose, which was suggested to be secondary to the suppression of the hypoglycaemic effect of endogenous IGF-I. However, the deletion of the *IGFBP-1* gene in mice did not show significant changes in glucose metabolism (Wheatcroft and Kearney, 2009). Studies have also shown that the expression of *IGFBP*-1 mRNA is affected by metabolic disturbances (Beale, 2013, Gu et al., 2014, Heald et al., 2001). Heald et al. (2001) observed an association between low circulating levels of *IGFBP-1* and the development of macrovascular disease and hypertension in T2D patients (Heald et al., 2001). It was recently observed that *IGFBP*-1 had lower concentrations in women with GDM as well as in the cord blood of their foetuses (Lappas, 2015). During pregnancy complicated by T1D, maternal plasma IGFBP-1 was found to be elevated more than two fold compared with normal pregnancy (Lappas, 2015). Another study has shown that cord blood *IGFBP-1* levels negatively correlate with birth weight and that IGFBP-1 levels are lower in pregnancies complicated by T2D and GDM in comparison to those measured in normal pregnancies (Hiden et al., 2009).

In healthy pregnancies not complicated by diabetes, IGF-1 is the most important growth factor *in utero* and is predominantly bound to IGFBP-1 which is a known regulator of foetal growth

and differentiation. Gestational diabetes is associated with increased birth weight and high rates of macrosomia (Lindsay et al., 2007). It is generally accepted that maternal hyperglycemia and foetal hyperinsulinemia are the main cause of this excessive growth, however, the role that IGFs and their binding proteins play in excessive foetal growth are less clear. The expression of *IGFBP-1* and its relative protein levels have been shown to be decreased in individuals with GDM (Lappas, 2015). Decreased placental expression levels of *IGFBP*-1 have been demonstrated in pregnancies with foetal growth restriction (FGR), however, whether these alterations are a causative factor of FGR or accompany other pathogenic mechanisms requires further investigation (Koutsaki et al., 2011). IGFBP-1 has not been extensively studied in the offspring of pregnant women, however studies examining IGFBP-1 concentrations in cord blood obtained from neonates exposed to a pregnancy complicated by diabetes have been inconsistent, with high and low concentrations found (Culler et al., 1996, Lindsay et al., 2007, Yang et al., 1996).

Several posttranslational changes modulate the affinity of IGFBP-1 for IGFs, particularly to IGF-I. These include phosphorylation/de-phosphorylation, proteolysis and polymerization and methylation (Gupta, 2015) and this is mentioned in a number of studies (Gibson et al., 1995, Shen et al., 2015).

4.1.2 *IGFBP-2*

Like *IGFBP-1, IGFBP-2* is a modulator of IGF-1 and IGF-2 bioavailability and plays an important role in the regulation of several cellular processes (Shen et al., 2015). IGFBP-2 is the most abundant IGFBP and is expressed in several tissues (Jones and Clemmons, 1995). Besides binding to IGFs in circulation, the IGF-regulatory activities of IGFBP-2 involve interactions with components of the extracellular matrix, cell surface proteoglycans and integrin receptors (Yau et al., 2015). IGFBP-2 exerts other key functions within the nucleus where it directly or indirectly promotes the transcription of other genes. All these activities of IGFBP-2, whether dependent or independent on IGFs, contribute to functional roles in growth and development (Yau et al., 2015).

IGFBP-2 has been previously linked to obesity and diabetes. Studies have shown that the administration of IGFBP-2 can be beneficial in improving metabolic responses (Yau et al., 2015). The association between IGFBP-2 levels and obesity are well established but are not yet well understood in terms of diabetes. In 2016, Kammel et al. (2016) observed a significant association between decreased *IGFBP*-2 expression and the development of obesity in adolescence. Allen et al. (2003) observed decreased plasma *IGFBP-2* levels in obese males and females (Allen et al., 2003). It is hypothesized that the reduced circulating levels of *IGFBP-2* in obese individuals are associated with an increased bioavailability of IGF-1, and hence stimulated growth (Kammel et al., 2016).

In terms of diabetes, the actions of *IGFBP-2* have been linked to insulin, although only in cases of hyperinsulemia where it appears to play a role in autocrine control in adipocytes (Aguirre et al., 2016). Lappas et al. (2015) found that low postnatal levels of *IGFBP-2* are a "significant risk factor for the development of T2D in women with a previous history of GDM". This study reports *IGFBP-2* as a potential biomarker for the prediction of T2D in women who developed GDM (Lappas, 2015). They also found that cord plasma levels of *IGFBP-2* were significantly and negatively correlated to the fasting glucose level in an OGTT when corrected for maternal BMI. It is not clear as to whether a decreased expression of IGFBP-2 may contribute to the pathophysiology of GDM or whether it is a result of the chronic insulin resistance of these patients (Retnakaran, 2016).

4.1.3 *IGFBP-6*

The data relating to the expression of *IGFBP*-6 and its association with diabetes are contradictory. In a study using diabetic rats, reduced levels of *IGFBP*-6 mRNA were observed when compared to healthy rats (Bergman et al., 2005). In 2012, *IGFBP*-6 serum levels were found to be significantly higher in patients with T1D (Bach, 2015c) however no associations with T2D or GDM have been reported. In an animal study using sheep, it was observed that placental weight increased between days 45 and 90 of gestation and this increase was accompanied by a reduction in *IGFBP*-6 expression. Lappas (2015) found that the maternal plasma levels of IGFBP-6 were significantly lower in obese individuals with NGT compared with

NGT non-obese and non-obese women with GDM. This finding suggests that reduced levels of IGFBP-6 may play a role in the development of obesity.

4.1.4 The possible implication of maternal *IGFBP-1* and *IGFBP-2* in abnormal foetal growth associated with gestational diabetes

Throughout pregnancy, the expression, circulating levels and covalent modifications of IGFBPs continuously change in the mother and may influence circulating IGF bioavailability (Sferruzzi-Perri et al., 2011). The maternal IGF system plays a vital role in foetal growth regulation via stimulation of extravillous trophoblast migration/invasion and facilitation of nutrient exchange through the promotion of growth and development of the placenta (Qiu et al., 2005). Decreased maternal pregravid insulin resistance coupled with an inadequate insulin response are the pathophysiological mechanisms underlying the development of gestational diabetes (Catalano et al., 2003). It is well known that IGFBP-1 acts to inhibit IGF function and is inversely related to insulin levels.

In my study, the blood samples taken from the mothers who develop gestational diabetes show a significant down regulation of *IGFBP-1* mRNA expression when compared to controls. IGFBP-2 and IGFBP-6 appear to have lower levels of mRNA expression in GDM cases when compared to controls, however, this difference does not reach levels of significance. The levels of *IGFBP-1* and *IGFBP-2* mRNA expression in maternal blood samples are significantly associated with maternal glucose levels at fasting and at 1-hr post OG (and at 2-hr post OG for *IGFBP-1* only). No correlation is observed between maternal glucose levels and *IGFBP-6* mRNA expression in maternal blood samples. There is a significant negative correlation between the levels of IGFBP-1 mRNA expression in maternal blood and foetal growth. There was no statistical association observed between maternal IGFBP mRNA expression and BMI. The significant under expression of *IGFBP-1* mRNA in women with GDM is not associated with promoter region methylation and no association was observed between maternal glucose levels and IGFBP methylation.

The changes in maternal carbohydrate metabolism (increased glucose and insulin levels) that result from the presence of gestational diabetes may lead to a decrease in maternal IGFBP-1

levels. It may be possible that high glucose levels cause a decrease in IGFBP-1 mRNA expression (Zhang et al., 2010). Given the role of IGFBP-1 in regulating IGF-1 bioavailability, the reduced IGFBP-1 levels in women diagnosed with GDM may be an important contributor to the development of insulin resistance through increased IGF-1 bioavailability. Therefore, we hypothesize that the significantly lower level of IGFBP-1 mRNA expression observed in women with GDM may be due to the inhibitory effect of high glucose levels that are characteristic of women with GDM. It is also possible that high insulin levels in women with GDM may also cause reduced expression of maternal IGFBPs (Gibson et al., 1996, Heald et al., 2001, Rajwani et al., 2012) but unfortunately, one of the limitations to this study was that the fasting insulin levels of the mothers were not measured and therefore an association between the levels of IGFBP-1 mRNA expression and insulin concentration could not be determined.

Our data have shown that the maternal IGFBP-1 gene expression levels are inversely correlated with birth weight, as was also observed in other studies (Verhaeghe et al., 1993, Whittaker et al., 1990). This data supports the hypothesis that low levels of maternal IGFBP-1 may lead to increased levels of bioavailable IGF peptides, thereby increasing IGF-induced foetal growth (Giudice et al., 1995). Alterations in circulating IGF and IGFBPs may alter birth weight and/or neonatal adiposity (Lappas, 2015). This may lead to a change in the ideal growth trajectories, resulting in the development of metabolic disorders later in life. Data from this study suggest that the level of *IGFBP-1* mRNA expression in the mother may be regulated by her glucose levels. Decreased expression of IGFBP-1, as observed in GDM in this and other studies, is likely a consequence of the presence of GDM-associated glycaemia, leading to greater bioavailable IGFs which in turn induce the characteristic macrosomia observed in neonates in GDM pregnancies. Although this is a possibility, we do not see a significant difference in the birthweight between the GDM and the control group.

The slightly lower (but statistically non-significant) level of IGFBP-2 expression in blood cells from GDM compared to non-GDM women and the lack of correlation between levels of IGFBP-2 mRNA expression and maternal BMI and foetal birth weight may indicate that maternal IGFBP-2 does not play a significant role in the outcome of foetal size in gestational diabetes.

However, in a study performed by Zhu et al. (2016), they found that the maternal plasma levels of IGFBP-2 decreased throughout pregnancies affected by GDM. They also observed a significant association between higher IGFBP-2 levels and a lower risk of developing GDM (Zhu et al., 2016). This finding together with the data from this study showing a trend towards lower IGFBP-2 expression in GDM women, lead us to hypothesise that IGFBP-2 is implicated to some extent in the development of GDM. It is possible that we don't see a significant under expression of this gene in our cohort due to small sample size.

In terms of maternal IGFBP-6, we do not observe any significant difference in the level of expression between the case and control groups. There were no significant correlations between the mRNA expression of this gene in maternal blood and other variables such as maternal glucose levels, maternal BMI and foetal birth weight. In the literature there is an established link between low maternal plasma IGFBP-6 levels and obesity (Hair et al., 2015, Lappas, 2015) whereas the link between maternal IGFBP-6 and gestational diabetes is not so well defined (Ferrero et al., 2012; Lappas et al., 2015). Therefore we can conclude that it is unlikely that the high glucose and insulin levels of the mother have any significant effect on IGFBP-6 mRNA expression.

4.1.5 Placental IGFBPs mRNA expression and possible implications for foetal growth

Gestational diabetes contributes to adverse foetal outcomes such as the development of insulin resistance *in utero* (Catalano et al., 2009) and large-for-gestational age (Ferraro et al., 2012) neonates. One of the many functions of the placenta is to produce a number of growth factors which may regulate the growth and the functions of the placenta in an autocrine or paracrine manner. The insulin-like growth factors (IGFs) have been shown to play a vital role in foetal growth and development (Baker et al., 1993, Liu et al., 1996). At the feto-maternal interface, *IGFBP-1* to *IGFBP-6* are predominantly expressed by the decidua during human pregnancy, with *IGFBP-1* being the most abundant (Hill et al., 1993). Additionally, *IGFBP-3*, *IGFBP-4* and *IGFBP-5* are expressed by the chorionic mesoderm (Han and Carter, 2000). Given IGFBP expression at the feto-maternal interface, IGFBPs may play a significant role in modulating IGF actions in the placenta, as well as having IGF-independent effects. Although the molecular mechanisms controlling normal foetal growth and development remain poorly

understood, it is thought that the placenta may play an important role through the secretion of placental hormones (Freemark, 2010). These hormones are secreted either into the umbilical circulation to directly affect foetal metabolism and growth, or into the maternal circulation to alter maternal metabolism and substrate availability for placental transfer (Sferuzzi-Perri et al., 2011).

The data presented in this study reveal that there is a significant down regulation of *IGFBP-1*, *IGFBP-2* and *IGFBP-6* mRNA expression in the placenta samples taken from infants born to women who developed gestational diabetes. The suppression of IGFBP-1 in the placenta exposed to gestational diabetes can be logically explained by hyperglycemia and raised maternal and foetal insulin levels, as insulin inhibits IGFBP-1 production (Loukovaara et al., 2005). The levels of *IGFBP-1*, *IGFBP-2* and *IGFBP-6* mRNA expression in the placenta were significantly negatively associated with maternal glucose levels at fasting and at 1-hr post OG (and at 2-hr post OG for *IGFBP-1* only). This finding may indicate that the reduced expression of these IGFBPs is a result of high maternal glucose and hyperinsulinaemia that is characteristic of gestational diabetes. Collectively, these data may suggest that high maternal glucose levels during pregnancy influence the bioavailability of IGFs indirectly through regulating IGFBPs, ultimately increasing foetal tissue and overall foetal somatic growth, resulting in larger than normal weight babies that is characteristic of GDM (Lappas, 2015). An alternative explanation is that the placental *IGFBP* expression might impact placental growth (and perhaps indirectly foetal growth). However, in this study we do not see a significant difference in the birthweight between the GDM and control groups.

IGFBP-1, IGFBP-2 and *IGFBP-6* mRNA expression levels are negatively correlated with foetal birth weight. This finding may implicate these IGFBPs in mechanisms underlying foetal growth. Reduced levels of these binding proteins in the placenta will result in an increase in free, unbound IGFs. These IGFs are then free to bind to their respective receptors, thus promoting foetal growth. It is possible for the free IGFs to cross the placenta, stimulating this effect in the foetal circulation, however, IGFs do not cross the placenta in very large quantities (Sferruzzi-Perri et al., 2011).

4.1.6 Maternal and placental IGFBPs promoter region methylation and foetal growth

In this study we did not observe significantly altered promoter region methylation in the blood of women with GDM for either IGFBP-1 or IGFBP-2. It appears that the significant decreased expression of *IGFBP-1* in the women with GDM is not associated with promoter region methylation. This supports our hypothesis that the reduced expression of IGFBP-1 mRNA is likely to be a result of inhibition due to high levels of glucose and/or insulin that is characteristic of GDM women.

In the placental tissue, the under expression of *IGFBP-1* and *IGFBP-2* mRNA is significantly correlated with hypermethylation at the promoter region of these genes. We therefore conclude that hypermethylation at the promoter region significantly reduces the expression of *IGFBP-1* and *IGFBP-2;* which could possibly have an effect on the bioavailability of certain growth factors (namely IGF-1) and consequently effect foetal growth. This hypothesis is supported by the observation of a significant correlation between foetal birth weight and the level of promoter region methylation in placental *IGFBP-1* and *IGFBP-2.* Our data suggests that an increase in maternal glucose levels will have an effect on the methylation and expression of these binding proteins in the placenta. This results in increased foetal growth and larger birth weight babies, as observed with GDM. The data suggest that DNA methylation may reduce expression of these binding proteins in the placenta but not in the mother. It may be essential that the placenta requires the ability to respond to prevailing nutrient levels and an epigenetic process would allow that, whilst in the adult linking growth to nutrients would be less essential than in the placenta and/or foetus. We conclude that the hypermethylation observed at the promoter region of these binding proteins in the placenta, may be a result of the presence of GDM and not an adaptive response by the foetus. The absence of *IGFBP-1* promoter methylation in the blood from women with GDM may suggest that any effect that GDM has on *IGFBP-1* should be reversible in the mothers after birth (as GDM is a transient form of the disease). Conversely, the presence of *IGFBP-1* promoter methylation in placental tissue obtained from GDM exposed foetuses suggests that the effect of GDM on *IGFBP-1* may potentially be manifested long term in the infants after birth.

4.2 The Pentose Phosphate Pathway (PPP)

Gestational diabetes mellitus is characterised by hyperglycemia that results from an insulin supply that is inadequate to overcome the rise of insulin resistance that occurs during pregnancy (Buchanan and Xiang, 2005). Glucose can have different metabolic fates once taken up into a cell. Glycolysis, the first step of which is the production of glucose-6-phosphate, is the main pathway of glucose metabolism, generating ATP and pyruvate. Alternative pathways for glucose-6-phosphate are into the PPP or to glycogen synthesis. The PPP (also referred to as the hexose monophosphate shunt) is a metabolic pathway parallel to glycolysis. It generates precursors for the synthesis of coenzymes, nucleotides, RNA and DNA (ribose-5-phosphate) and nicotinamide adenosine dinucleotide phosphate (NADPH). The shift of glucose to this pathway tends to occur in organs with intense growth or with high biosynthesis demands (Michalek et al., 2011). The activity of the PPP is regulated according to the immediate metabolic situation and the needs of the organism; if not needed, the products of PPP are readily converted to glycolytic intermediates and oxidized. Shunting of accumulated cytosolic glycolytic intermediates into the PPP supposedly unburdens glycolysis and quantitatively limits processing of glycolytic intermediates into harmful metabolic products. In this way, the PPP represents a potentially 'protective' mechanism against hyperglycaemia-induced damage (Pacal et al., 2011).

There are two distinct phases in the PPP. The first is the oxidative phase, in which NADPH is generated, and the second is the non-oxidative synthesis of 5-carbon sugars. Although the primary role of the PPP is anabolic rather than catabolic, it does still involve the oxidation of glucose. Both the process of glycolysis and the PPP appear to be highly conserved amongst species and have a very ancient evolutionary origin suggesting importance in function (Court et al., 2015, Stincone et al., 2015). One of the major uses of NADPH in the cell is to prevent oxidative stress. Oxidative stress can be defined as any "disturbance in the balance of antioxidants and pro-oxidants in favour of the later due to different factors such as aging, drug actions and toxicity and inflammation" (Rahal et al., 2014). Oxidative stress causes healthy cells of the body to lose their function and structure from oxygen free radical damage and when the antioxidant level is limited, this type of damage can become debilitating (Asmat et

al., 2016). Damage to DNA, proteins, and other macromolecules due to oxygen free radicals has been implicated in the pathogenesis of a wide variety of diseases, including diabetes (Asmat et al., 2016).

Nicotinamide adenine dinucleotide phosphate (NADPH) reduces glutathione via glutathione reductase, which converts reactive H_2O_2 into H_2O by glutathione peroxidase. If NADPH were absent, the H_2O_2 would be converted to hydroxyl free radicals which can cause irreversible cell damage. Cells generate a large amount of NADPH through the PPP to use in the reduction of glutathione. There are two routes for the disposal of H_2O_2: by catalase and by glutathione peroxidase (Gaetani et al., 1996). It is known that a defect in the catalase route for the disposal of peroxide can predispose individuals to developing diabetes. Therefore it is possible that it a defect in the glutathione peroxidase route (due to aberrant NADPH production) could have similar effects. It has been known for more than two decades that reactive oxygen species can cause diabetes in rodents (Bondeva and Wolf, 2014) and it is known that increased oxidative damage is found in both T1D and T2D (Giacco and Brownlee, 2010). Therefore, it has been suggested that oxidative stress is a significant cause of diabetes in human populations(Lappas et al., 2011, Matough et al., 2012). On the other hand, a number of studies have concluded that diabetes causes oxidative stress (Inoguchi et al., 2000) which may in turn cause the vascular and microvascular-complications that are characteristic of poorly controlled diabetes.

Pathological pregnancies, including GDM, are associated with higher levels of oxidative stress which is the result of the overproduction of free radicals and/or a defect in the antioxidant defences (Lappas et al., 2011). In GDM pregnancies, glucose tolerance and metabolism as well as insulin resistance are altered, and although the pathophysiologic mechanisms underlying these changes are not completely understood, they are accompanied by oxidative stress (Zhu et al., 2015).

In this study we see the significant under expression of five key enzymes that play a role in the PPP namely *G6PD*, *TKT*, *PGLS, DCXR* and *ALDOA* in the blood of women who develop GDM. Two of these genes, *G6PD* and *TKT*, were also found to be significantly under expressed in

exposed placenta. However, only *G6PD* was found to have significant promoter region methylation in both the blood of women who develop GDM as well as exposed placenta.

Hyperglycaemia is the major feature of GDM. In theory, glucose is normally metabolized through the process of glycolysis but in the presence of abnormally high levels, some glucose is shunted to the PPP in an attempt to unburden glycolysis. In this way, the PPP acts as a protective mechanism to prevent the processing of glycolytic intermediates into toxic end products. One would therefore expect an increase in the action of the enzymes of this pathway in subjects with hyperglycaemia. However, in this study we observed a decrease in the expression of the enzymes in women who develop GDM (*G6PD* and *TKT* in exposed placenta as well). The reduced expression of key enzymes in the PPP (*G6PD*, *TKT*, *ALDOA, PGLS* and *DCXR)* should lead to lower activity of the pathway. This will result in less NAPDH being produced in both the women and the placenta, putting them at risk for increased oxidative stress.

Only a few studies have investigated the PPP in the human placenta. Glucose oxidation via the tricarboxylic acid (TCA) cycle and PPP, measured through the ratio of CO_2 production from [1-C]-glucose and [6-C]-glucose, have suggested that oxidation via the TCA cycle is higher than via the PPP in term placental slices, whereas in younger placentas (6-20 weeks) oxidation by the PPP is greater (Brekke et al., 2012). Different proportions of glucose utilisation via glycolysis and the TCA cycle or the PPP were found in another research study, but the PPP also accounted for higher portions of glucose utilisation (10 %) in early pregnancy (6-10 weeks) compared to term placentas (5 %)(Bertoldi Franco, 2015). In human trophoblast cultures from term placentas, the PPP accounted for less than 1% of the total glucose metabolised (Brekke et al., 2012). It is unknown whether glucose partitioning into the PPP increases with increasing glucose supply and whether a PPP functioning at a reduced rate is an adaptive mechanism in response to exposure to an adverse environment.

4.2.1 A significant decrease in the expression of the gene (*G6PD*) encoding the rate determining enzyme of the PPP

The *G6PD* gene is an X-linked gene that maps to the Xq28 and is the rate-controlling enzyme of the PPP pathway. It is allosterically stimulated by $NADP^+$ and strongly inhibited by NADPH (Patra and Hay, 2014). This cytoplasmic enzyme catalyses the rate-limiting step in the oxidative branch of the PPP that generates the first molecule of NADPH, therefore its expression and activity are tightly regulated (Jiang et al., 2014, Jin et al., 2014). A cell's defence mechanism against oxidative damage is highly dependent on the activity of this enzyme because it is the main source of NADPH. The activity of *G6PD* is also post-translationally regulated by cytoplasmic deacetylase *SIRT2*. The *SIRT2*-mediated deacetylation and activation of *G6PD* stimulates the oxidative branch of the PPP to supply cytosolic NADPH to counteract oxidative damage (Xu et al., 2016b, Zhu et al., 2015). A study has shown that *SIRT2* was not found to have significant differential expression in women with GDM and exposed placenta when compared to controls (Gui et al., 2015). It has been reported in a study by Zhang et al. (2010), that an increase in the concentration of maternal glucose resulted in the inhibition of *G6PD* and consequently a decrease in *NADPH* levels (Zhang et al., 2010). The data from this study supports this finding. We also observed a significant negative association between G6PD mRNA expression levels and maternal glucose levels. Hyperglycemia is known to elevate oxidative stress and also increase the activation of the hexosamine biosynthetic pathway, a secondary pathway of glucose metabolism. This pathway uses fructose-6-phosphate derived from glycolysis to produce glucosamine-6-phosphate which is a competitive inhibitor of *G6PD*. The inhibition or decrease in *G6PD* leads to decreased *NAPDH* concentrations and elevated oxidative stress. Numerous observations have demonstrated highly significant decreases in *G6PD* activity due to hyperglycemia or diabetes in liver, kidney, brain, endothelium, red blood cells and other cells and tissues (Stanton, 2012). Decreased *G6PD* activity has been observed in cells and tissues from diabetic animals. For example, neutrophils exposed to an increase in glucose levels had impaired neutrophil function associated with decreased *G6PD* function. This suggests that the glucose mediated decrease in *G6PD* led to a decrease in *NADPH* production that was needed for *NAPDH* oxidase activity in neutrophils (Perner et al., 2003, Stanton, 2012).

Human *G6PD* deficiency has mainly been studied in the context of the associated haemolysis or protection from malaria (Ouattara et al., 2014, Peters and Van Noorden, 2009, Valencia et al., 2016). However, some studies have reported an association between the activity of G6PD and other diseases. For example, a study performed in the Middle East demonstrated an increased frequency of diabetes in individual's who were *G6PD* deficient (Pinna et al., 2013). Epidemiological data suggest that G6PD deficiency may be a risk factor for diabetes (Carette et al., 2011). Mutations in the *G6PD* gene results in protein variants with varying levels of enzyme activity accounting for a wide spectrum of clinical and biochemical phenotypes – although many individuals don't display symptoms of *G6PD* deficiency. Deficiency of G6PD, also known as favism, can also be caused by an X-linked recessive genetic condition that predisposes those affected to developing haemolysis (Gaskin et al., 2011). Deficiency in *G6PD* is commonly found in people of Mediterranean and African origin due to its anti-malaria effects (Beutler et al., 2007).

In our study we observe a significant negative correlation between the maternal glucose levels (at fasting and 1-hr post OGTT) and G6PD mRNA expression in both the mother and the placenta. Studies have found that high glucose levels have an inhibitory effect on the expression of G6PD (Zhang et al., 2000). Modest changes in G6PD activity have significant effects on cell growth and cell death in a variety of cell types (Tian et al., 1998), highlighting the importance of adequate functioning of G6PD. An association between high glucose levels and a decrease in *G6PD* expression and activity in human islets has been observed (Zhang et al., 2010). A high glucose level has been reported to suppress *G6PD* activity in endothelial cells, kidney, liver, and red blood cells resulting in oxidative damage, cellular dysfunction, and organ damage (Cheng et al., 2000, Xu et al., 2005). Gestational diabetes results in maternal hyperglycemia, which will untimely increase the flow of glucose from the mother, across the placenta, to the foetus. The foetus will be exposed to an *in utero* environment of hyperglycemia and hyperinsulemia. The significant reduction in the expression of *G6PD* in both blood from the mother and the placenta may be due to the inhibitory effects of high glucose levels.

Our data also showed a significant negative correlation between the promoter region methylation of the *G6PD* gene and G6PD gene expression in both the placenta and the maternal blood cells. It appears from these observations that the significant decrease in *G6PD* expression in a hyperglycemic environment may be due to increased promoter region methylation. In a recent study carried out by Xu et al. (2016), an association between low *G6PD* expression and increased promoter region methylation had been observed. This association was significant in females but not in males (Xu et al., 2016a). Our data did show a significant positive correlation between the promoter region methylation of G6PD in the blood and placenta with maternal glucose levels (at fasting and at 1-hr post OG). This may indicate that the expression of *G6PD* is inhibited by high glucose levels through an epigenetic mechanism, which is DNA methylation. The finding of reduced *G6PD* expression in the mother and placenta with hyperglycemia is potentially important. There seems to be a strong rationale for believing that high maternal glucose leads to methylation of the *G6PD* promoter, which in turn leads to decreased *G6PD* expression. Since *G6PD* expression levels may increase oxidative stress, this is likely a pathological rather than adaptive response.

The same levels of G6PD promoter region methylation are observed in mothers and their fetuses and therefore, it becomes difficult to determine whether these methylation patterns were inherited or if they were directly caused by the hyperglycemic environment or if they were an adaptive response in both mother and foetus to the high glucose levels. From an adaptive perspective, the down-regulation of *G6PD* does not seem like an advantageous adaption for the survival of the developing foetus in a hyperglycemic environment. It has been extensively reported that *G6PD* is the principal source of NADPH and is critical for the defence against oxidative stress and a decrease in the expression of *G6PD* would result in a decrease in the production of NADPH. Given the evidence from the literature it is more likely that the high levels of glucose associated with GDM cause the reduction in the expression of *G6PD* in both maternal blood and the placenta. Therefore, we suggest that the significantly reduced expression of G6PD we observe in exposed placenta and maternal blood cells is due to the high glucose levels.

High glucose levels inhibit G6PD and therefore, in GDM, it may be possible that the PPP will be inhibited by increasing glucose metabolism by glycolysis. This will cause NADPH to drop (due to low G6PD activity) and oxygen free radical production to increase (due to higher flux through glycolysis). Both these events will contribute to increased oxidative stress. Furthermore, due to increased flux through glycolysis, the glycolytic pathway will become saturated leading to a plateau in glucose metabolism causing glucose levels to rise. Thus, the inhibition of G6PD by hyperglycaemia will indirectly worsen the hyperglycaemic state. The alternative hypothesis, for which there is less evidence, is that low G6PD activity is part of the aetiological pathway for GDM and diabetes. Low G6PD protein levels would push glucose into the glycolytic pathway increasing free radical output whilst NADPH would fall increasing oxidative stress. The glycolytic pathway may become saturated reducing glucose metabolism and causing blood glucose levels to increase. The high methylation levels in the GDM subjects may cause the reduced G6PD activity and one could hypothesise that this hypermethylation is due to other upstream factors including gene polymorphisms related to methylation at the G6PD locus. However, evidence from the literature to support this is sparse. One would assume that subjects with G6PD deficiency, which is quite common in African populations, may develop diabetes. Polymorphisms in the G6PD gene causing low levels of expression are quite common but none of them have been associated with diabetes in GWAS. GWAS studies cannot provide information about epigenetic effects, so the absence of a GWAS association between polymorphisms in *G6PD* and diabetes does not mean that there is no link between G6PD and epigenetic effects in GDM. Furthermore, there are studies associating diabetes with G6PD deficiency (Heymann et al., 2012, Pinna et al., 2013, Saeed et al., 1985) but none of these relate G6PD gene variants to diabetes but only measure G6PD activity. Therefore, these results could just as easily be interpreted as showing that diabetes causes lowered G6PD activity, as vice versa.

Interestingly, studies have shown that in GDM the level of insulin output is lower than in mothers without GDM at each level of insulin resistance (Megia et al., 2008). Thus, in GDM the ß-cells fail to produce sufficient insulin in response to the rise in insulin resistance during pregnancy. It may be possible that the G6PD deficiency we see in GDM women could be

involved in this process, with excess oxidative stress causing ß-cell dysfunction in the GDM mothers. Thus the increase in glucose during GDM could lead to low G6PD mRNA expression and higher oxidative stress across the mother and the foetus. High glucose may also reduce G6PD in the ß-cells leading to cell death and reduced insulin output (Stanton, 2012). Thus, G6PD deficiency may be a result of GDM but it may also enhance the pathological processes leading to eventual diabetes. These data suggest that G6PD deficiency on its own is not enough to cause diabetes but requires the presence of hyperglycaemia and insulin resistance before it will contribute to disease progression.

4.2.2 A significant reduction in the expression of the reversible rate-limiting enzyme, Transketolase, of the non-oxidative branch of the PPP

In mammals, transketolase (TKT) connects the PPP to glycolysis, feeding excess sugar phosphates into the main carbohydrate metabolic pathways. Functional TKT is also necessary for the production of *NADPH* which, as discussed above, is necessary to counteract oxidative stress.

In this study, we observed a significant decrease in the expression of TKT mRNA in the women who develop GDM. This down regulation of the TKT gene in the blood samples was negatively correlated to maternal fasting glucose levels. These data indicate that high glucose levels in the mother are associated with decreased TKT expression in the blood. The promoter region of the gene was not significantly hypermethylated in the GDM cases and also there was no correlation between TKT promoter region methylation and maternal glucose levels. Therefore we can conclude that promoter region methylation is not the cause of the down regulation of TKT that we observe in GDM. One explanation could be that due to a decrease in the expression of G6PD, genes encoding enzymes further downstream in the pathway (*PGLS*, *DCXR*, *ALDOA*, *TKT*) may have a lower activity and expression due to a lower level of substrate. Thus, the decreased expression of *G6PD* will cause a so called “ripple effect” which will lead to a sequential decrease in the concentration of the downstream metabolites and their associated enzymes.

In summary, our results suggest that high glucose levels that are characteristic of GDM, may result in the suppression of G6PD mRNA expression in both maternal blood and placental

tissue. We hypothesize that this suppression is mediated through increased DNA methylation of the *G6PD* gene. Our results also suggest that decreased IGFBP-1 expression in the placenta and in the mother in a GDM pregnancy, is implicated in the increased foetal growth which is characteristic of GDM.

4.3 The limitations of this study

One of the most apparent limitations of this study was the use of whole blood to identify gene expression and methylation changes in GDM cases and controls. Whole blood is made up of many different cell types, each of which may have a subtly different epigenetic pattern and gene expression profile. Performing DNA methylation analysis on a DNA sample that was obtained from a multicellular tissue will result in the average DNA methylation profile of that tissue. The varying levels of methylation in each cell type and the varying proportion in which each cell type is present in the tissue may mask significant methylation differences in a specific cell type that may be critical to disease in question.

The question regarding which tissue is the most ideal in epigenetic studies is also of concern. The DNA methylation and gene expression patterns are tissue specific which imposes an issue when studying a disease which is known to act through perturbation of a specific cell or tissue type. The analysis of epigenetic patterns, in particular DNA methylation, of a surrogate non-target tissue may not be informative. The most appropriate tissue for studying GDM is the pancreas, but obtaining samples from this tissue is not possible. For ethical and practical reasons, the collection of human DNA samples must be as non-invasive as possible. Saliva, buccal swabs, blood samples or biopsies are the most common sources of DNA and RNA. DNA from peripheral blood and placental biopsies were used in this study, with the limitations recognised. More work needs to be done to determine whether these tissues are good surrogates for the specific diabetes relevant target tissue.

The major limitation of this study is the small sample size. Determining the ideal sample size is important to ensure that you would be able to confidently observe a true effect or to

determine if there is sufficient power to detect a meaningful difference. The small sample size of the study limits the power of detection, therefore it is necessary to validate these finding in a larger cohort. The sample size should be increased to increase the statistical power of detecting true correlations in gene expression between the women with GDM and controls (Joehanes et al., 2013, Jones et al., 2015, Rapaport et al., 2013, Venet et al., 2011). To confirm that these changes are generally relevant would require a larger sample, as would the detection of more subtle gene expression shifts in GDM. However, although the samples size was small we did detect meaningful changes in gene expression and methylation in GDM-relevant pathways that can be investigated in greater detail in future studies which use a much larger sample.

Confounding is a key limitation of the work presented in this study, with specific reference to, but not restricted to, BMI. The control group had a BMI that was on average 7kg/m^2 higher than the case group. This is a marked difference and the results presented here could potentially and importantly be confounded by BMI. The lack of replication of the data presented in this study is also considered a limitation.

4.4. Strengths of the study

The biggest strength of this study is the homogeneous sample group of patients that was studied. All patients were of the same ethnicity and similar age. Gene expression and methylation patterns observed in the patients from this study are more likely to underlie disease aetiology than be confounded by heterogeneity arising from different ethnicities, ages and BMI of the patients. In addition, we only studied cases with female offspring. Another strength is that the methods used to detect (RNA-seq) and validate (qPCR) the significant gene expression differences between the cases and control groups were different, but the outcome was highly correlated. This promotes confidence in the results obtained. Despite the small sample size of the study, we did observe significant variation between the case and control groups in terms of gene expression and methylation. Also, we present strong correlations between the many variables analysed in this study. Another strength of this study is that we

investigated GDM in an under-studied population that has a high prevalence of obesity and type 2 diabetes and is therefore at high risk for GDM.

4.5 Future work

The search for candidate genes that predispose women to GDM is currently a very active field of research, and more and more genes associated with the development of the disease are being discovered. Before any further studies are done, there is a need for additional exploration to determine whether the results could be explained by confounding or by chance. A genome wide DNA methylation and transcriptomic analysis should be carried out on a larger cohort.

In the future, it would be imperative to conduct an hypothesis driven study to further explore the role of the PPP in GDM and whether it is mediated by an essentially epigenetic mechanism, rather than genetic variation. Alternatively there may be a genetic susceptibility that drives the epigenetic modulation in the presence of an adverse environment. Future studies should focus on *G6PD* and *NADPH* and their possible role in the development of adult onset diseases in foetuses exposed to hyperglycemia and these studies should be carried out in a larger cohort. It would also be of interest to include cord blood as an additional tissue to study as this will give a clearer idea of what is happening in the developing foetus when placental G6PD is suppressed and oxidative stress increases in the placenta. Likewise, studying the neonate, infant, child, adolescent and adult through their life time is enormously important in determining the trajectories of change that leads to adult onset disease. It may also be of interest to look at promoter region methylation in the other genes that make up the PPP, and NADPH levels in a variety of tissues and developmental time points.

Future studies should focus on the precise mechanisms by which high maternal glucose levels observed in the GDM cases may cause a suppression in the expression of *G6PD* in maternal blood and placental tissue. We hypothesize that it may be due to promoter region methylation because our data indicated a significant correlation between expression of *G6PD* and

promoter region methylation in both tissues studied. These studies could potentially focus on other epigenetic mechanisms such as histone modifications and gene-gene interactions through techniques such as CHiP-seq or ATAC-seq.

It would be important to focus a larger study on the IGFBPs and other components of the IGF-axis and examine their relative mRNA expression and protein levels in placenta and offspring exposed to GDM. The IGF-axis may be implicated in glucose homeostasis but its longitudinal profile across gestation in relation to the development of GDM is largely unknown. A critical unresolved question is whether or not the IGF-axis contributes to the pathophysiology of GDM. A key question in future studies would be whether DNA methylation at these loci is affected by early life exposures (such as the nutritional insult GDM imposes) and if so whether these methylation changes persist through early life and into adulthood. To investigate this DNA methylation analysis of the placental tissue would have to be undertaken at multiple time points including *in utero*, neonatal and later in life. This would allow the stability of epigenetic marks to be considered in light of the emergence of programmed phenotypic changes.

Future studies could also include a more detailed testing and documentation of participant's phenotypic data as this will allow robust and precise correlations between the clinical data and gene expression patterns. As mentioned, protein levels would be of interest in these future studies, therefore, since gene expression changes identified in the transcriptome do not always translate directly into changes in protein production, another validation approach that should accompany future studies is immunohistochemistry. This way, more accurate correlations between gene expression and phenotype (based on clinical data) can be generated.

It would also be interesting to look at *G6PD* expression and methylation patterns in newly diagnosed type 2 diabetic and control subjects to determine if the observations in the current study are specific to GDM. The relationship between *G6PD* deficiency and T2D is not clear. Therefore a study should be undertaken where newly diagnosed type 2 diabetics are compared against a healthy control group in terms of G6PD activity levels in blood cells and the frequency of *G6PD* polymorphisms. A longitudinal study would also be useful to determine

whether G6PD deficiency and/or *G6PD* polymorphisms predict the future development of T2D in a population with a high risk of diabetes.

Chapter FIVE

5. CONCLUSIONS

The enzyme *G6PD,* is at the nexus of many important metabolic pathways and we are only at the beginning of understanding *G6PD*, *NADPH* and their interrelationships with cellular systems. A full understanding of *G6PD*, its role as a metabolic nexus for many cellular systems and how it is regulated should provide critical insights into many intracellular processes and disease mechanisms (Stanton, 2012).

The high circulating glucose levels that are characteristic of GDM may have a suppressive effect on the expression of *G6PD* mRNA in the blood of affected women, however, further work would need to be carried out in order to determine a possible causative link. The developing foetus is also exposed to high levels of glucose, which cross the placenta via diffusion. This may result in the suppression of the expression of placental *G6PD* mRNA. This reduction of *G6PD*, the rate determining step of the pentose phosphate pathway, will affect the rate at which this pathway functions. The reaction catalyzed by *G6PD* gives rise to NADPH which plays a role in reducing oxidative stress by counteracting oxidative damage. Oxidative stress, which is known to play a role in the pathogenesis of diabetes and other adult disease, will increase in the presence or reduced levels of NADPH. We hypothesize that the mechanism by which high glucose suppresses *G6PD* expression is via DNA methylation. It is known that GDM is a transient condition, therefore after birth, the women's glucose levels will return to normal but the methylation and reduced expression of *G6PD* may persist. The suppression of *G6PD* and hence the reduction of NADPH may continue and increased cellular oxidative stress ensues. There is convincing experimental evidence that demonstrates that an increase in ROS increases in diabetes and oxidative stress is associated with the onset of diabetes (Matough et al., 2012). This may explain why women who develop GDM have a higher risk of developing T2D and other metabolic disorders later in life. The down-regulation of *G6PD* may also lead to the overloading of the glycolytic pathway with glucose, leading to reduced glucose uptake and

increased levels of blood glucose. It is possible that this, in combination with the insulin resistance of pregnancy uncovers an underlying problem in glucose metabolism and insulin secretory capacity resulting in GDM. However, with the increase in insulin sensitivity after birth, the metabolic pathways may be less stressed and euglycaemia is maintained but the underlying pathology may worsen over time leading to the development of T2D.

It is possible that *G6PD* is an important component of the placental protection system. In this capacity, the optimal regulation of the gene encoding this enzyme is important for the health of the foetus. In terms of the developing foetus exposed to the adverse hyperglycemic *in utero* environment, the suppression of G6PD and NADPH will result in increased oxidative stress. This increase might interfere with the optimal functioning of the placenta. Oxidative stress has been observed in the placenta in preeclampsia and diabetes in association with altered placental function (Hanson and Mair, 2014, Myatt and Cui, 2004).

This study provides evidence of aberrant regulation of *IGFBPs* (*IGFBP-1*, *IGFBP-2* and *IGFBP-6*) in the context of GDM exposure. Our data suggests that the suppression of gene expression of these three binding proteins in the placenta may partially account for the higher birth weight of the neonates born to women with GDM. We conclude that the foetal hyperglycemia and/or hyperinsulemia caused by maternal GDM, inhibits the expression of *IGFBP-1* and *IGFBP-2* ultimately affecting foetal growth through regulating the bioavailaibity of IGFs.

It is now widely accepted that the development of chronic diseases in adulthood may have their origins in the womb. The development of GDM can exert both short and long term adverse effects on the health of the developing foetus (Brenseke et al., 2013). The challenges at present are to "identify common mechanisms and pathways involved in different perinatal malnutrition paradigms, deciphering physiological and/or pathological roles of specific nutrients, and to determine which components of the maternal diet may be best modified to optimize maternal health, placental integrity, birth outcome, and lifelong health of the offspring" (Brenseke, 2015).

The role of early adverse life exposures in the developmental programming of adult onset diseases such as diabetes is well documented and it seems that DNA methylation may play a

key role in this process (Vaiserman et al., 2013; Lewis et al., 2014). However, it still remains unclear as to the precise moment during development where exposure to an adverse environment will have the greatest effect. Animal studies demonstrate that exposure to an adverse *in utero* environment programs the pattern of DNA methylation which persists throughout adulthood ultimately affecting gene expression levels and therefore the phenotype (Lee et al., 2015; Williams et al., 2014; Vaiserman et al., 2015). There has also been a suggestion that the pubertal period may be the critical window in development and that changes established at this particular stage may persist throughout the life course (Jasik and Lustig, 2008) and result in the development of disease. The Newcastle thousand family study suggests that influences in childhood are the most important determinants of disease risk in adulthood (Pearce et al., 2012). These studies highlight the importance of not only focusing on one particular developmental stage but rather assessing the role of epigenetic variation at many developmental stages. In addition, the possible role of genetic susceptibility in triggering DNA methylation changes or other epigenetic modulations in the presence of an adverse environment remains to be explored.

The identification of genes linked to GDM will contribute to our understanding of the pathophysiology of the disease and to the development of prevention strategies. Identifying individuals who have a genetic predisposition to developing GDM may improve prevention of T2D through targeted interventions. These data provide an interesting starting point in the investigation of developmental programming in children exposed to GDM. Demonstrating a causal link between gene specific differential DNA methylation and gene expression is a promising target for future research.

REFERENCES

AGROGIANNIS, G. D., SIFAKIS, S., PATSOURIS, E. S. & KONSTANTINIDOU, A. E. 2014. Insulin-like growth factors in embryonic and fetal growth and skeletal development (Review). *Mol Med Rep,* 10**,** 579-84.

AGUIRRE, G. A., DE ITA, J. R., DE LA GARZA, R. G. & CASTILLA-CORTAZAR, I. 2016. Insulin-like growth factor-1 deficiency and metabolic syndrome. *J Transl Med,* 14**,** 3.

AIRD, D., ROSS, M. G., CHEN, W. S., DANIELSSON, M., FENNELL, T., RUSS, C., JAFFE, D. B., NUSBAUM, C. & GNIRKE, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol,* 12**,** R18.

ALDASOUQI, S. A., SOLOMON, D. J., BOKHARI, S. A., KHAN, P. M., MUNEERA, S. & GOSSAIN, V. V. 2008. Glycohemoglobin A1c: A promising screening tool in gestational diabetes mellitus. *Int J Diabetes Dev Ctries,* 28**,** 121-4.

ALFARADHI, M. Z. & OZANNE, S. E. 2011. Developmental programming in response to maternal overnutrition. *Front Genet,* 2**,** 27.

ALLEN, N. E., APPLEBY, P. N., KAAKS, R., RINALDI, S., DAVEY, G. K. & KEY, T. J. 2003. Lifestyle determinants of serum insulin-like growth-factor-I (IGF-I), C-peptide and hormone binding protein levels in British women. *Cancer Causes Control,* 14**,** 65-74.

ANGUEIRA, A. R., LUDVIK, A. E., REDDY, T. E., WICKSTEED, B., LOWE, W. L., JR. & LAYDEN, B. T. 2015. New insights into gestational glucose metabolism: lessons learned from 21st century approaches. *Diabetes,* 64**,** 327-34.

ARORA, D., ARORA, R., SANGTHONG, S., LEELAPORN, W. & SANGRATANATHONGCHAI, J. 2013. Universal screening of gestational diabetes mellitus: prevalence and diagnostic value of clinical risk factors. *J Med Assoc Thai,* 96**,** 266-71.

ASMAT, U., ABAD, K. & ISMAIL, K. 2016. Diabetes mellitus and oxidative stress-A concise review. *Saudi Pharm J,* 24**,** 547-553.

ATAGUBA, J. E., AKAZILI, J. & MCINTYRE, D. 2011. Socioeconomic-related health inequality in South Africa: evidence from General Household Surveys. *Int J Equity Health,* 10**,** 48.

AZEVEDO, M. & ALLA, S. 2008. Diabetes in sub-saharan Africa: kenya, mali, mozambique, Nigeria, South Africa and zambia. *Int J Diabetes Dev Ctries,* 28**,** 101-8.

BACH, L. A. 2015a. Insulin-Like Growth Factor Binding Proteins--an Update. *Pediatr Endocrinol Rev,* 13**,** 521-30.

BACH, L. A. 2015b. Insulin-like growth factor binding proteins 4-6. *Best Pract Res Clin Endocrinol Metab,* 29**,** 713-22.

BACH, L. A. 2015c. Recent insights into the actions of IGFBP-6. *J Cell Commun Signal,* 9**,** 189-200.

BAE, C. H., KIM, J. S., SONG, S. Y., KIM, Y. W., PARK, S. Y. & KIM, Y. D. 2013. Insulin-like growth factor-1 induces MUC8 and MUC5B expression via ERK1 and p38 MAPK in human airway epithelial cells. *Biochem Biophys Res Commun,* 430**,** 683-8.

BAKER, J., LIU, J. P., ROBERTSON, E. J. & EFSTRATIADIS, A. 1993. Role of insulin-like growth factors in embryonic and postnatal growth. *Cell,* 75**,** 73-82.

BANCROFT, K., TUFFNELL, D. J., MASON, G. C., ROGERSON, L. J. & MANSFIELD, M. 2000. A randomised controlled pilot study of the management of gestational impaired glucose tolerance. *BJOG,* 107**,** 959-63.

BARFIELD, R. T., ALMLI, L. M., KILARU, V., SMITH, A. K., MERCER, K. B., DUNCAN, R., KLENGEL, T., MEHTA, D., BINDER, E. B., EPSTEIN, M. P., RESSLER, K. J. & CONNEELY, K. N. 2014. Accounting for population stratification in DNA methylation studies. *Genet Epidemiol,* 38**,** 231-41.

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.

BAROUKI, R., GLUCKMAN, P. D., GRANDJEAN, P., HANSON, M. & HEINDEL, J. J. 2012. Developmental origins of non-communicable disease: implications for research and public health. *Environ Health,* 11**,** 42.

BEAGLEHOLE, R. & YACH, D. 2003. Globalisation and the prevention and control of non-communicable disease: the neglected chronic diseases of adults. *Lancet,* 362**,** 903-8.

BEALE, E. G. 2013. Insulin signaling and insulin resistance. *J Investig Med,* 61**,** 11-4.

BEEN, L. F., RALHAN, S., WANDER, G. S., MEHRA, N. K., SINGH, J., MULVIHILL, J. J., ASTON, C. E. & SANGHERA, D. K. 2011. Variants in KCNQ1 increase type II diabetes susceptibility in South Asians: a study of 3,310 subjects from India and the US. *BMC Med Genet,* 12**,** 18.

BELL, J. T., PAI, A. A., PICKRELL, J. K., GAFFNEY, D. J., PIQUE-REGI, R., DEGNER, J. F., GILAD, Y. & PRITCHARD, J. K. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol,* 12**,** R10.

BELLINGER, L., LILLEY, C. & LANGLEY-EVANS, S. C. 2004. Prenatal exposure to a maternal low-protein diet programmes a preference for high-fat foods in the young adult rat. *Br J Nutr,* 92**,** 513-20.

BEN-HAROUSH, A., YOGEV, Y. & HOD, M. 2004. Epidemiology of gestational diabetes mellitus and its association with Type 2 diabetes. *Diabet Med,* 21**,** 103-13.

BERGER, H. & SERMER, M. 2009. Counterpoint: Selective screening for gestational diabetes mellitus. *Diabetes Care,* 32**,** 1352-4.

BERGMAN, P. B., MORAVSKI, C. J., EDMONDSON, S. R., RUSSO, V. C., BACH, L. A., WILKINSON-BERKA, J. L. & WERTHER, G. A. 2005. Expression of the IGF system in normal and diabetic transgenic (mRen-2)27 rat eye. *Invest Ophthalmol Vis Sci,* 46**,** 2708-15.

BERNASKO, J. 2016. Gestational Diabetes Screening: The International Association of the Diabetes and Pregnancy Study Groups Compared With Carpenter-Coustan Screening. *Obstet Gynecol,* 127**,** 964-5.

BERTOLDI FRANCO, C. 2015. *Effects of glucose and fatty acids interactions on glucose metabolism pathways in human trophoblasts* MPhil, Australian National University.

BEUTLER, E., DUPARC, S. & GROUP, G. P. D. W. 2007. Glucose-6-phosphate dehydrogenase deficiency and antimalarial drug development. *Am J Trop Med Hyg,* 77**,** 779-89.

BIERNE, H., HAMON, M. & COSSART, P. 2012. Epigenetics and bacterial infections. *Cold Spring Harb Perspect Med,* 2**,** a010272.

BIRD, A. 2002. DNA methylation patterns and epigenetic memory. *Genes Dev,* 16**,** 6-21.

BLEWITT, M. & WHITELAW, E. 2013. The use of mouse models to study epigenetics. *Cold Spring Harb Perspect Biol,* 5**,** a017939.

BOKS, M. P., DERKS, E. M., WEISENBERGER, D. J., STRENGMAN, E., JANSON, E., SOMMER, I. E., KAHN, R. S. & OPHOFF, R. A. 2009. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One,* 4**,** e6767.

BOLLATI, V. & BACCARELLI, A. 2010. Environmental epigenetics. *Heredity (Edinb),* 105**,** 105-12.

BONDEVA, T. & WOLF, G. 2014. Reactive oxygen species in diabetic nephropathy: friend or foe? *Nephrol Dial Transplant,* 29**,** 1998-2003.

BORGHOL, N., SUDERMAN, M., MCARDLE, W., RACINE, A., HALLETT, M., PEMBREY, M., HERTZMAN, C., POWER, C. & SZYF, M. 2012. Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol,* 41**,** 62-74.

BOTTALICO, J. N. 2007. Recurrent gestational diabetes: risk factors, diagnosis, management, and implications. *Semin Perinatol,* 31**,** 176-84.

BOUCHARD, L. 2013. Epigenetics and fetal metabolic programming: a call for integrated research on larger cohorts. *Diabetes,* 62**,** 1026-8.

BOUCHARD, L., HIVERT, M. F., GUAY, S. P., ST-PIERRE, J., PERRON, P. & BRISSON, D. 2012. Placental adiponectin gene DNA methylation levels are associated with mothers' blood glucose concentration. *Diabetes,* 61**,** 1272-80.

BOUCHARD, L., THIBAULT, S., GUAY, S. P., SANTURE, M., MONPETIT, A., ST-PIERRE, J., PERRON, P. & BRISSON, D. 2010. Leptin gene epigenetic adaptation to impaired glucose metabolism during pregnancy. *Diabetes Care,* 33**,** 2436-41.

BREKKE, E. M., WALLS, A. B., SCHOUSBOE, A., WAAGEPETERSEN, H. S. & SONNEWALD, U. 2012. Quantitative importance of the pentose phosphate pathway determined by incorporation of 13C from [2-13C]- and [3-13C]glucose into TCA cycle intermediates and neurotransmitter amino acids in functionally intact neurons. *J Cereb Blood Flow Metab,* 32**,** 1788-99.

BRENSEKE, B., PRATER, M. R., BAHAMONDE, J. & GUTIERREZ, J. C. 2013. Current thoughts on maternal nutrition and fetal programming of the metabolic syndrome. *J Pregnancy,* 2013**,** 368461.

BUCHANAN, T. A. 2001. Pancreatic B-cell defects in gestational diabetes: implications for the pathogenesis and prevention of type 2 diabetes. *J Clin Endocrinol Metab,* 86**,** 989-93.

BUCHANAN, T. A. & XIANG, A. H. 2005. Gestational diabetes mellitus. *J Clin Invest,* 115**,** 485-91.

BUCHANAN, T. A., XIANG, A. H. & PAGE, K. A. 2012. Gestational diabetes mellitus: risks and management during and after pregnancy. *Nat Rev Endocrinol,* 8**,** 639-49.

BUTTE, N. F. 2000. Carbohydrate and lipid metabolism in pregnancy: normal compared with gestational diabetes mellitus. *Am J Clin Nutr,* 71**,** 1256S-61S.

CARETTE, C., DUBOIS-LAFORGUE, D., GAUTIER, J. F. & TIMSIT, J. 2011. Diabetes mellitus and glucose-6-phosphate dehydrogenase deficiency: from one crisis to another. *Diabetes Metab,* 37**,** 79-82.

CARPENTER, M. W. & COUSTAN, D. R. 1982. Criteria for screening tests for gestational diabetes. *Am J Obstet Gynecol,* 144**,** 768-73.

CATALANO, P. M., KIRWAN, J. P., HAUGEL-DE MOUZON, S. & KING, J. 2003. Gestational diabetes and insulin resistance: role in short- and long-term implications for mother and fetus. *J Nutr,* 133**,** 1674S-1683S.

CATALANO, P. M., PRESLEY, L., MINIUM, J. & HAUGUEL-DE MOUZON, S. 2009. Fetuses of obese mothers develop insulin resistance in utero. *Diabetes Care,* 32**,** 1076-80.

CAZALY, E., CHARLESWORTH, J., DICKINSON, J. L. & HOLLOWAY, A. F. 2015. Genetic Determinants of Epigenetic Patterns: Providing Insight into Disease. *Mol Med,* 21**,** 400-9.

CECCARELLI, M., BARTHEL, F. P., MALTA, T. M., SABEDOT, T. S., SALAMA, S. R., MURRAY, B. A., MOROZOVA, O., NEWTON, Y., RADENBAUGH, A., PAGNOTTA, S. M., ANJUM, S., WANG, J., MANYAM, G., ZOPPOLI, P., LING, S., RAO, A. A., GRIFFORD, M., CHERNIACK, A. D., ZHANG, H., POISSON, L., CARLOTTI, C. G., JR., TIRAPELLI, D. P., RAO, A., MIKKELSEN, T., LAU, C. C., YUNG, W. K., RABADAN, R., HUSE, J., BRAT, D. J., LEHMAN, N. L., BARNHOLTZ-SLOAN, J. S., ZHENG, S., HESS, K., RAO, G., MEYERSON, M., BEROUKHIM, R., COOPER, L., AKBANI, R., WRENSCH, M., HAUSSLER, D., ALDAPE, K. D., LAIRD, P. W., GUTMANN, D. H., NETWORK, T. R., NOUSHMEHR, H., IAVARONE, A. & VERHAAK, R. G. 2016. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell,* 164**,** 550-63.

CHARLES, M. A., DELPIERRE, C. & BREANT, B. 2016. [Developmental origin of health and adult diseases (DOHaD): evolution of a concept over three decades]. *Med Sci (Paris),* 32**,** 15-20.

CHEN, M. & ZHANG, L. 2011. Epigenetic mechanisms in developmental programming of adult disease. *Drug Discov Today,* 16**,** 1007-18.

CHENG, M. L., HO, H. Y., LIANG, C. M., CHOU, Y. H., STERN, A., LU, F. J. & CHIU, D. T. 2000. Cellular glucose-6-phosphate dehydrogenase (G6PD) status modulates the effects of nitric oxide (NO) on human foreskin fibroblasts. *FEBS Lett,* 475**,** 257-62.

CHO, H., MU, J., KIM, J. K., THORVALDSEN, J. L., CHU, Q., CRENSHAW, E. B., 3RD, KAESTNER, K. H., BARTOLOMEI, M. S., SHULMAN, G. I. & BIRNBAUM, M. J. 2001. Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta). *Science,* 292**,** 1728-31.

CHRISTENSEN, B. C., HOUSEMAN, E. A., MARSIT, C. J., ZHENG, S., WRENSCH, M. R., WIEMELS, J. L., NELSON, H. H., KARAGAS, M. R., PADBURY, J. F., BUENO, R., SUGARBAKER, D. J., YEH, R. F., WIENCKE, J. K. & KELSEY, K. T. 2009. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet,* 5**,** e1000602.

CITRO, A., CANTARELLI, E. & PIEMONTI, L. 2015a. The CXCR1/2 Pathway: Involvement in Diabetes Pathophysiology and Potential Target for T1D Interventions. *Curr Diab Rep,* 15**,** 68.

CITRO, A., VALLE, A., CANTARELLI, E., MERCALLI, A., PELLEGRINI, S., LIBERATI, D., DAFFONCHIO, L., KASTSIUCHENKA, O., RUFFINI, P. A., BATTAGLIA, M., ALLEGRETTI, M. & PIEMONTI, L. 2015b. CXCR1/2 inhibition blocks and reverses type 1 diabetes in mice. *Diabetes,* 64**,** 1329-40.

CLAESSON, R., ABERG, A. & MARSAL, K. 2007. Abnormal fetal growth is associated with gestational diabetes mellitus later in life: population-based register study. *Acta Obstet Gynecol Scand,* 86**,** 652-6.

CLOONAN, N. & GRIMMOND, S. M. 2008. Transcriptome content and dynamics at single-nucleotide resolution. *Genome Biol,* 9**,** 234.

COCQUET, J., CHONG, A., ZHANG, G. & VEITIA, R. A. 2006. Reverse transcriptase template switching and false alternative transcripts. *Genomics,* 88**,** 127-31.

COLBERG, S. R., CASTORINO, K. & JOVANOVIC, L. 2013. Prescribing physical activity to prevent and manage gestational diabetes. *World J Diabetes,* 4**,** 256-62.

CONERLY, M. & GRADY, W. M. 2010. Insights into the role of DNA methylation in disease through the use of mouse models. *Dis Model Mech,* 3**,** 290-7.

CONESA, A., MADRIGAL, P., TARAZONA, S., GOMEZ-CABRERO, D., CERVERA, A., MCPHERSON, A., SZCZESNIAK, M. W., GAFFNEY, D. J., ELO, L. L., ZHANG, X. & MORTAZAVI, A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol,* 17**,** 13.

COOKSON, W., LIANG, L., ABECASIS, G., MOFFATT, M. & LATHROP, M. 2009. Mapping complex disease traits with global gene expression. *Nat Rev Genet,* 10**,** 184-94.

COSTA, V., ANGELINI, C., DE FEIS, I. & CICCODICOLA, A. 2010. Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol,* 2010**,** 853916.

COSTA, V., APRILE, M., ESPOSITO, R. & CICCODICOLA, A. 2013. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *Eur J Hum Genet,* 21**,** 134-42.

COURT, S. J., WACLAW, B. & ALLEN, R. J. 2015. Lower glycolysis carries a higher flux than any biochemically possible alternative. *Nat Commun,* 6**,** 8427.

CRUME, T. L., OGDEN, L., DANIELS, S., HAMMAN, R. F., NORRIS, J. M. & DABELEA, D. 2011. The impact of in utero exposure to diabetes on childhood body mass index growth trajectories: the EPOCH study. *J Pediatr,* 158**,** 941-6.

CULLER, F. L., TUNG, R. F., JANSONS, R. A. & MOSIER, H. D. 1996. Growth promoting peptides in diabetic and non-diabetic pregnancy: interactions with trophoblastic receptors and serum carrier proteins. *J Pediatr Endocrinol Metab,* 9**,** 21-9.

DAKOVIC, D., COLIC, M., CAKIC, S., MILEUSNIC, I., HAJDUKOVIC, Z. & STAMATOVIC, N. 2013. Salivary interleukin-8 levels in children suffering from type 1 diabetes mellitus. *J Clin Pediatr Dent,* 37**,** 377-80.

DAXINGER, L. & WHITELAW, E. 2010. Transgenerational epigenetic inheritance: more questions than answers. *Genome Res,* 20**,** 1623-8.

DAVEY SMITH, G & EBRAHIM. 2004. Mendelian randomization: prospects, potentials and limitations. I*nt J Epidemiol*. 33(1) 30 - 42.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of transcription. *Genes Dev,* 25**,** 1010-22.

DEIERLEIN, A. L., SIEGA-RIZ, A. M., HERRING, A. H., ADAIR, L. S. & DANIELS, J. L. 2012. Gestational weight gain and predicted changes in offspring anthropometrics between early infancy and 3 years. *Pediatr Obes,* 7**,** 134-42.

DOLINOY, D. C. 2008. The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Nutr Rev,* 66 Suppl 1**,** S7-11.

DOLINOY, D. C. & JIRTLE, R. L. 2008. Environmental epigenomics in human health and disease. *Environ Mol Mutagen,* 49**,** 4-8.

DONADI, E., CEZAR, N., EVANGELISTA, A., XAVIER, D., ASSIS, A., ARNS, T., 2014. Gestational diabetes mellitus transcriptome analysis revels induction of genes related to the major histocompatibility complex. *Hum Immunol* 75:pg 122.

DURAN, A., SAENZ, S., TORREJON, M. J., BORDIU, E., DEL VALLE, L., GALINDO, M., PEREZ, N., HERRAIZ, M. A., IZQUIERDO, N., RUBIO, M. A., RUNKLE, I., PEREZ-FERRE, N., CUSIHUALLPA, I., JIMENEZ, S., GARCIA DE LA TORRE, N., FERNANDEZ, M. D., MONTANEZ, C., FAMILIAR, C. & CALLE-PASCUAL, A. L. 2014. Introduction of IADPSG criteria for the screening and diagnosis of gestational diabetes mellitus results in improved pregnancy outcomes at a lower cost in a large cohort of pregnant women: the St. Carlos Gestational Diabetes Study. *Diabetes Care,* 37**,** 2442-50.

DURNWALD, C. 2015. Gestational diabetes: Linking epidemiology, excessive gestational weight gain, adverse pregnancy outcomes, and future metabolic syndrome. *Semin Perinatol,* 39**,** 254-8.

DURON, E., FUNALOT, B., BRUNEL, N., COSTE, J., QUINQUIS, L., VIOLLET, C., BELMIN, J., JOUANNY, P., PASQUIER, F., TRELUYER, J. M., EPELBAUM, J., LE BOUC, Y. & HANON, O. 2012. Insulin-like growth factor-I and insulin-like growth factor binding protein-3 in Alzheimer's disease. *J Clin Endocrinol Metab,* 97**,** 4673-81.

EADES, C. E., STYLES, M., LEESE, G. P., CHEYNE, H. & EVANS, J. M. 2015. Progression from gestational diabetes to type 2 diabetes in one region of Scotland: an observational follow-up study. *BMC Pregnancy Childbirth,* 15**,** 11.

EDGREN, H., MURUMAGI, A., KANGASPESKA, S., NICORICI, D., HONGISTO, V., KLEIVI, K., RYE, I. H., NYBERG, S., WOLF, M., BORRESEN-DALE, A. L. & KALLIONIEMI, O. 2011. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol,* 12**,** R6.

EL-MAARRI, O., BECKER, T., JUNEN, J., MANZOOR, S. S., DIAZ-LACAVA, A., SCHWAAB, R., WIENKER, T. & OLDENBURG, J. 2007. Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. *Hum Genet,* 122**,** 505-14.

EL HAJJ, N., PLIUSHCH, G., SCHNEIDER, E., DITTRICH, M., MULLER, T., KORENKOV, M., ARETZ, M., ZECHNER, U., LEHNEN, H. & HAAF, T. 2013. Metabolic programming of MEST DNA methylation by intrauterine exposure to gestational diabetes mellitus. *Diabetes,* 62**,** 1320-8.

EL HAJJ, N., SCHNEIDER, E., LEHNEN, H. & HAAF, T. 2014a. Epigenetics and life-long consequences of an adverse nutritional and diabetic intrauterine environment. *Reproduction,* 148**,** R111-R120.

EL HAJJ, N., SCHNEIDER, E., LEHNEN, H. & HAAF, T. 2014b. Epigenetics and life-long consequences of an adverse nutritional and diabetic intrauterine environment. *Reproduction,* 148**,** R111-20.

ELNOUR, A. A., EL MUGAMMAR, I. T., JABER, T., REVEL, T. & MCELNAY, J. C. 2008. Pharmaceutical care of patients with gestational diabetes mellitus. *J Eval Clin Pract,* 14**,** 131-40.

ENQUOBAHRIE, D. A., WILLIAMS, M. A., QIU, C., MELLER, M. & SORENSEN, T. K. 2009. Global placental gene expression in gestational diabetes mellitus. *Am J Obstet Gynecol,* 200**,** 206 e1-13.

EVANGELISTA, A., COLLARES, C., XAVIER, D., MAREDO, C., MANOEL-CAETANO, FS., RASSI, DM. 2014. Intergrative analysis of the transcriptome profiles profiles observed in type 1, type 2 and gestational diabetes mellitus revelas the role of inflammation. *BMC Med Genom* 7(28).

FAN, S. & ZHANG, X. 2009. CpG island methylation pattern in different human tissues and its correlation with gene expression. *Biochem Biophys Res Commun,* 383**,** 421-5.

FEIL, R & FRAGA, M. 2012. Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 13(2): 97-109

FERNANDEZ, A. F., ASSENOV, Y., MARTIN-SUBERO, J. I., BALINT, B., SIEBERT, R., TANIGUCHI, H., YAMAMOTO, H., HIDALGO, M., TAN, A. C., GALM, O., FERRER, I., SANCHEZ-CESPEDES, M., VILLANUEVA, A., CARMONA, J., SANCHEZ-MUT, J. V., BERDASCO, M., MORENO, V., CAPELLA, G., MONK, D., BALLESTAR, E., ROPERO, S., MARTINEZ, R., SANCHEZ-CARBAYO, M., PROSPER, F., AGIRRE, X., FRAGA, M. F., GRANA, O., PEREZ-JURADO, L., MORA, J., PUIG, S., PRAT, J., BADIMON, L., PUCA, A. A., MELTZER, S. J., LENGAUER, T., BRIDGEWATER, J., BOCK, C. & ESTELLER, M. 2012. A DNA methylation fingerprint of 1628 human samples. *Genome Res,* 22**,** 407-19.

FERRARO, Z. M., QIU, Q., GRUSLIN, A. & ADAMO, K. B. 2012. Characterization of the insulin-like growth factor axis in term pregnancies complicated by maternal obesity. *Hum Reprod,* 27**,** 2467-75.

FINER, S., MATHEWS, C., LOWE, R., SMART, M., HILLMAN, S., FOO, L., SINHA, A., WILLIAMS, D., RAKYAN, V. K. & HITMAN, G. A. 2015. Maternal gestational diabetes is associated with genome-wide DNA methylation variation in placenta and cord blood of exposed offspring. *Hum Mol Genet,* 24**,** 3021-9.

FORBES, B. E., MCCARTHY, P. & NORTON, R. S. 2012. Insulin-like growth factor binding proteins: a structural perspective. *Front Endocrinol (Lausanne),* 3**,** 38.

FRASER, H. B., LAM, L. L., NEUMANN, S. M. & KOBOR, M. S. 2012. Population-specificity of human DNA methylation. *Genome Biol,* 13**,** R8.

FRAZEE, A. C., SABUNCIYAN, S., HANSEN, K. D., IRIZARRY, R. A. & LEEK, J. T. 2014. Differential expression analysis of RNA-seq data at single-base resolution. *Biostatistics,* 15**,** 413-26.

FREEMARK, M. 2010. Placental hormones and the control of fetal growth. *J Clin Endocrinol Metab,* 95**,** 2054-7.

FREINKEL, N., COCKROFT, D. L., LEWIS, N. J., GORMAN, L., AKAZAWA, S., PHILLIPS, L. S. & SHAMBAUGH, G. E., 3RD 1986. The 1986 McCollum award lecture. Fuel-mediated teratogenesis during early organogenesis: the effects of increased concentrations of glucose, ketones, or somatomedin inhibitor during rat embryo culture. *Am J Clin Nutr,* 44**,** 986-95.

FUNNELL, M. M., BROWN, T. L., CHILDS, B. P., HAAS, L. B., HOSEY, G. M., JENSEN, B., MARYNIUK, M., PEYROT, M., PIETTE, J. D., READER, D., SIMINERIO, L. M., WEINGER, K. & WEISS, M. A. 2007. National standards for diabetes self-management education. *Diabetes Educ,* 33**,** 599-600, 602-4, 606 passim.

GABBE, S. G. 1986. Gestational diabetes mellitus. *N Engl J Med,* 315**,** 1025-6.

GAETANI, G. F., FERRARIS, A. M., ROLFO, M., MANGERINI, R., ARENA, S. & KIRKMAN, H. N. 1996. Predominant role of catalase in the disposal of hydrogen peroxide within human erythrocytes. *Blood,* 87**,** 1595-9.

GAUNT, T. R., SHIHAB, H. A., HEMANI, G., MIN, J. L., WOODWARD, G., LYTTLETON, O., ZHENG, J., DUGGIRALA, A., MCARDLE, W. L., HO, K., RING, S. M., EVANS, D. M., DAVEY SMITH, G. & RELTON, C. L. 2016. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol,* 17**,** 61.

GEORGE, S., ROCHFORD, J. J., WOLFRUM, C., GRAY, S. L., SCHINNER, S., WILSON, J. C., SOOS, M. A., MURGATROYD, P. R., WILLIAMS, R. M., ACERINI, C. L., DUNGER, D. B., BARFORD, D., UMPLEBY, A. M., WAREHAM, N. J., DAVIES, H. A., SCHAFER, A. J., STOFFEL, M., O'RAHILLY, S. & BARROSO, I. 2004. A family with severe insulin resistance and diabetes due to a mutation in AKT2. *Science,* 304**,** 1325-8.

GHOSH, S., DENT, R., HARPER, M., GORMAN, S., STUART, JS., MCPHERSON, R. 2010. Gene expression profiling in whole blood identifies distinct biological pathways associated with obesity. *BMC Med Genom* 3:56

GIACCO, F. & BROWNLEE, M. 2010. Oxidative stress and diabetic complications. *Circ Res,* 107**,** 1058-70.

GIBBS, J. R., VAN DER BRUG, M. P., HERNANDEZ, D. G., TRAYNOR, B. J., NALLS, M. A., LAI, S. L., AREPALLI, S., DILLMAN, A., RAFFERTY, I. P., TRONCOSO, J., JOHNSON, R., ZIELKE, H. R., FERRUCCI, L., LONGO, D. L., COOKSON, M. R. & SINGLETON, A. B. 2010. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet,* 6**,** e1000952.

GIBSON, J. M., WESTWOOD, M., CROSBY, S. R., GORDON, C., HOLLY, J. M., FRASER, W., ANDERSON, C., WHITE, A. & YOUNG, R. J. 1995. Choice of treatment affects plasma levels of insulin-like growth factor-binding protein-1 in noninsulin-dependent diabetes mellitus. *J Clin Endocrinol Metab,* 80**,** 1369-75.

GIBSON, J. M., WESTWOOD, M., YOUNG, R. J. & WHITE, A. 1996. Reduced insulin-like growth factor binding protein-1 (IGFBP-1) levels correlate with increased cardiovascular risk in non-insulin dependent diabetes mellitus (NIDDM). *J Clin Endocrinol Metab,* 81**,** 860-3.

GILMARTIN, A. B., URAL, S. H. & REPKE, J. T. 2008. Gestational diabetes mellitus. *Rev Obstet Gynecol,* 1**,** 129-34.

GIUDICE, L. C., DE ZEGHER, F., GARGOSKY, S. E., DSUPIN, B. A., DE LAS FUENTES, L., CRYSTAL, R. A., HINTZ, R. L. & ROSENFELD, R. G. 1995. Insulin-like growth factors and their binding proteins in the term and preterm human fetus and neonate with normal and extremes of intrauterine growth. *J Clin Endocrinol Metab,* 80**,** 1548-55.

GLUCKMAN, P. D. & HANSON, M. A. 2004. The developmental origins of the metabolic syndrome. *Trends Endocrinol Metab,* 15**,** 183-7.

GODFREY, K. M., SHEPPARD, A., GLUCKMAN, P. D., LILLYCROP, K. A., BURDGE, G. C., MCLEAN, C., RODFORD, J., SLATER-JEFFERIES, J. L., GARRATT, E., CROZIER, S. R., EMERALD, B. S., GALE, C. R., INSKIP, H. M., COOPER, C. & HANSON, M. A. 2011. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. *Diabetes,* 60**,** 1528-34.

GONCALVES, I., BENGTSSON, E., COLHOUN, H. M., SHORE, A. C., PALOMBO, C., NATALI, A., EDSFELDT, A., DUNER, P., FREDRIKSON, G. N., BJORKBACKA, H., OSTLING, G., AIZAWA, K., CASANOVA, F., PERSSON, M., GOODING, K., STRAIN, D., KHAN, F., LOOKER, H. C., ADAMS, F., BELCH, J., PINNOLI, S., VENTURI, E., KOZAKOVA, M., GAN, L. M., SCHNECKE, V., NILSSON, J. & CONSORTIUM, S. 2015. Elevated Plasma Levels of MMP-12 Are Associated With Atherosclerotic Burden and Symptomatic Cardiovascular Disease in Subjects With Type 2 Diabetes. *Arterioscler Thromb Vasc Biol,* 35**,** 1723-31.

GONZALEZ-PARRA, S., ROSATO, R. R., CHOWEN, J. A., ARGENTE, J., GROFFEN, C., DITS, N. & DROP, S. L. 2002. Inverse correlation between insulin-like growth factor (IGF)-binding protein-5 and IGF-I and II during postnatal development of the anterior pituitary gland. *Horm Res,* 57**,** 10-4.

GU, T., FALHAMMAR, H., GU, H. F. & BRISMAR, K. 2014. Epigenetic analyses of the insulin-like growth factor binding protein 1 gene in type 1 diabetes and diabetic nephropathy. *Clin Epigenetics,* 6**,** 10.

GUI, L. S., ZHANG, Y. R., LIU, G. Y. & ZAN, L. S. 2015. Expression of the SIRT2 gene and its relationship with body size traits in Qinchuan cattle (Bos taurus). *Int J Mol Sci,* 16**,** 2458-71.

GUPTA, M. B. 2015. The role and regulation of IGFBP-1 phosphorylation in fetal growth restriction. *J Cell Commun Signal,* 9**,** 111-23.

GUPTA, S., DEOSKAR, A., GUPTA, P. & JAIN, S. 2015. Serum insulin-like growth factor-1 levels in females and males in different cervical vertebral maturation stages. *Dental Press J Orthod,* 20**,** 68-75.

HAIR, B. Y., TROESTER, M. A., EDMISTON, S. N., PARRISH, E. A., ROBINSON, W. R., WU, M. C., OLSHAN, A. F., SWIFT-SCANLAN, T. & CONWAY, K. 2015. Body mass index is associated with gene methylation in estrogen receptor-positive breast tumors. *Cancer Epidemiol Biomarkers Prev,* 24**,** 580-6.

HALES, C. N. & BARKER, D. J. 1992. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia,* 35**,** 595-601.

HALIM, A. A., SZITA, N. & BAGANZ, F. 2013. Characterization and multi-step transketolase-omega-transaminase bioconversions in an immobilized enzyme microreactor (IEMR) with packed tube. *J Biotechnol,* 168**,** 567-75.

HALL, E., VOLKOV, P., DAYEH, T., ESGUERRA, J. L., SALO, S., ELIASSON, L., RONN, T., BACOS, K. & LING, C. 2014. Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biol,* 15**,** 522.

HALL, V., THOMSEN, R. W., HENRIKSEN, O. & LOHSE, N. 2011. Diabetes in Sub Saharan Africa 1999-2011: epidemiology and public health implications. A systematic review. *BMC Public Health,* 11**,** 564.

HAN, V. K. & CARTER, A. M. 2000. Spatial and temporal patterns of expression of messenger RNA for insulin-like growth factors and their binding proteins in the placenta of man and laboratory animals. *Placenta,* 21**,** 289-305.

HANDY, D. E., CASTRO, R. & LOSCALZO, J. 2011. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. *Circulation,* 123**,** 2145-56.

HANNUM, G., GUINNEY, J., ZHAO, L., ZHANG, L., HUGHES, G., SADDA, S., KLOTZLE, B., BIBIKOVA, M., FAN, J. B., GAO, Y., DECONDE, R., CHEN, M., RAJAPAKSE, I., FRIEND, S., IDEKER, T. & ZHANG, K. 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell,* 49**,** 359-67.

HANSON, K. K. & MAIR, G. R. 2014. Stress granules and Plasmodium liver stage infection. *Biol Open,* 3**,** 103-7.

HANSON, M., GLUCKMAN, P., BIER, D., CHALLIS, J., FLEMING, T., FORRESTER, T., GODFREY, K., NESTEL, P. & YAJNIK, C. 2004. Report on the 2nd World Congress on Fetal Origins of Adult Disease, Brighton, U.K., June 7-10, 2003. *Pediatr Res,* 55**,** 894-7.

HANSON, M. A. & GLUCKMAN, P. D. 2014. Early developmental conditioning of later health and disease: physiology or pathophysiology? *Physiol Rev,* 94**,** 1027-76.

HANSON, M. A. & SKINNER, M. K. 2016. Developmental origins of epigenetic transgenerational inheritance. *Environ Epigenet,* 2.

HAUROGNE, K., PAVLOVIC, M., ROGNIAUX, H., BACH, J. M. & LIEUBEAU, B. 2015. Type 1 Diabetes Prone NOD Mice Have Diminished Cxcr1 mRNA Expression in Polymorphonuclear Neutrophils and CD4+ T Lymphocytes. *PLoS One,* 10**,** e0134365.

HEALD, A. H., CRUICKSHANK, J. K., RISTE, L. K., CADE, J. E., ANDERSON, S., GREENHALGH, A., SAMPAYO, J., TAYLOR, W., FRASER, W., WHITE, A. & GIBSON, J. M. 2001. Close relation of fasting insulin-like growth factor binding protein-1 (IGFBP-1) with glucose tolerance and cardiovascular risk in two populations. *Diabetologia,* 44**,** 333-9.

HEARD, E. & MARTIENSSEN, R. A. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell,* 157**,** 95-109.

HEIJMANS, B. T., TOBI, E. W., STEIN, A. D., PUTTER, H., BLAUW, G. J., SUSSER, E. S., SLAGBOOM, P. E. & LUMEY, L. H. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A,* 105**,** 17046-9.

HEINDEL, J. J. & VANDENBERG, L. N. 2015. Developmental origins of health and disease: a paradigm for understanding disease cause and prevention. *Curr Opin Pediatr,* 27**,** 248-53.

HEYMANN, A. D., COHEN, Y. & CHODICK, G. 2012. Glucose-6-phosphate dehydrogenase deficiency and type 2 diabetes. *Diabetes Care,* 35**,** e58.

HEYN, H., MORAN, S., HERNANDO-HERRAEZ, I., SAYOLS, S., GOMEZ, A., SANDOVAL, J., MONK, D., HATA, K., MARQUES-BONET, T., WANG, L. & ESTELLER, M. 2013. DNA methylation contributes to natural human variation. *Genome Res,* 23**,** 1363-72.

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**,** 60-8.

HILL, D. J., CLEMMONS, D. R., RILEY, S. C., BASSETT, N. & CHALLIS, J. R. 1993. Immunohistochemical localization of insulin-like growth factors (IGFs) and IGF binding proteins -1, -2 and -3 in human placenta and fetal membranes. *Placenta,* 14**,** 1-12.

HILL, V. K., RICKETTS, C., BIECHE, I., VACHER, S., GENTLE, D., LEWIS, C., MAHER, E. R. & LATIF, F. 2011. Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity. *Cancer Res,* 71**,** 2988-99.

HILLS, F. A., ILES, R. K. & SULLIVAN, M. H. 2013. Differential proteolysis of insulin-like growth factor binding protein-1 (IGFBP-1) in pregnancy. *J Perinat Med,* 41**,** 241-9.

HIRASAWA, R., CHIBA, H., KANEDA, M., TAJIMA, S., LI, E., JAENISCH, R. & SASAKI, H. 2008. Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev,* 22**,** 1607-16.

HO, S. M., JOHNSON, A., TARAPORE, P., JANAKIRAM, V., ZHANG, X. & LEUNG, Y. K. 2012. Environmental epigenetics and its implication on disease risk and health outcomes. *ILAR J,* 53**,** 289-305.

HOBBS, A. & RAMSAY, M. 2015. Epigenetics and the burden of noncommunicable disease: a paucity of research in Africa. *Epigenomics,* 7**,** 627-39.

HOLMES, R. P., HOLLY, J. M. & SOOTHILL, P. W. 2000. Maternal insulin-like growth factor binding protein-1, body mass index, and fetal growth. *Arch Dis Child Fetal Neonatal Ed,* 82**,** F113-7.

HONG, Y., KIM, W. J., BANG, C. Y., LEE, J. C. & OH, Y. M. 2016. Identification of Alternative Splicing and Fusion Transcripts in Non-Small Cell Lung Cancer by RNA Sequencing. *Tuberc Respir Dis (Seoul),* 79**,** 85-90.

HORVATH, K., KOCH, K., JEITLER, K., MATYAS, E., BENDER, R., BASTIAN, H., LANGE, S. & SIEBENHOFER, A. 2010. Effects of treatment in women with gestational diabetes mellitus: systematic review and meta-analysis. *BMJ,* 340**,** c1395.

HOUDE, A. A., GUAY, S. P., DESGAGNE, V., HIVERT, M. F., BAILLARGEON, J. P., ST-PIERRE, J., PERRON, P., GAUDET, D., BRISSON, D. & BOUCHARD, L. 2013. Adaptations of placental and cord blood ABCA1 DNA methylation profile to maternal metabolic status. *Epigenetics,* 8**,** 1289-302.

HOUDE, A. A., ST-PIERRE, J., HIVERT, M. F., BAILLARGEON, J. P., PERRON, P., GAUDET, D., BRISSON, D. & BOUCHARD, L. 2014. Placental lipoprotein lipase DNA methylation levels are associated with gestational diabetes mellitus and maternal and cord blood lipid profiles. *J Dev Orig Health Dis,* 5**,** 132-41.

HUANG, J., CHEN, Z. H., REN, C. M., WANG, D. X., YUAN, S. X., WU, Q. X., CHEN, Q. Z., ZENG, Y. H., SHAO, Y., LI, Y., WU, K., YU, Y., SUN, W. J. & HE, B. C. 2015. Antiproliferation effect of evodiamine in human colon cancer cells is associated with IGF-1/HIF-1alpha downregulation. *Oncol Rep,* 34**,** 3203-11.

HULT, M., TORNHAMMAR, P., UEDA, P., CHIMA, C., BONAMY, A. K., OZUMBA, B. & NORMAN, M. 2010. Hypertension, diabetes and overweight: looming legacies of the Biafran famine. *PLoS One,* 5**,** e13582.

INOGUCHI, T., LI, P., UMEDA, F., YU, H. Y., KAKIMOTO, M., IMAMURA, M., AOKI, T., ETOH, T., HASHIMOTO, T., NARUSE, M., SANO, H., UTSUMI, H. & NAWATA, H. 2000. High glucose level and free fatty acid stimulate reactive oxygen species production through protein kinase C--dependent activation of NAD(P)H oxidase in cultured vascular cells. *Diabetes,* 49**,** 1939-45.

JAFFE, A. E. & IRIZARRY, R. A. 2014. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol,* 15**,** R31.

JAMISON, D. T., SHAHID-SALLES, S. A., JAMISON, J., LAWN, J. E. & ZUPAN, J. 2006. Incorporating Deaths Near the Time of Birth into Estimates of the Global Burden of Disease. *In:* LOPEZ, A. D., MATHERS, C. D., EZZATI, M., JAMISON, D. T. & MURRAY, C. J. L. (eds.) *Global Burden of Disease and Risk Factors.* Washington (DC).

JASIK, C. B. & LUSTIG, R. H. 2008. Adolescent obesity and puberty: the "perfect storm". *Ann N Y Acad Sci,* 1135**,** 265-79.

JAWAD, F. & EJAZ, K. 2016. Gestational diabetes mellitus in South Asia: Epidemiology. *J Pak Med Assoc,* 66**,** S5-7.

JIANG, P., DU, W. & WU, M. 2014. Regulation of the pentose phosphate pathway in cancer. *Protein Cell,* 5**,** 592-602.

JIN, B., LI, Y. & ROBERTSON, K. D. 2011. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer,* 2**,** 607-17.

JIN, L., ZHANG, H., CHEN, L., YANG, C., YANG, S., JIANG, W. & GU, Y. 2014. Combined overexpression of genes involved in pentose phosphate pathway enables enhanced D-xylose utilization by Clostridium acetobutylicum. *J Biotechnol,* 173**,** 7-9.

JIRTLE, R. L. & SKINNER, M. K. 2007. Environmental epigenomics and disease susceptibility. *Nat Rev Genet,* 8**,** 253-62.

JOEHANES, R., YING, S., HUAN, T., JOHNSON, A. D., RAGHAVACHARI, N., WANG, R., LIU, P., WOODHOUSE, K. A., SEN, S. K., TANRIVERDI, K., COURCHESNE, P., FREEDMAN, J. E., O'DONNELL, C. J., LEVY, D. & MUNSON, P. J. 2013. Gene expression signatures of coronary heart disease. *Arterioscler Thromb Vasc Biol,* 33**,** 1418-26.

JONES, A. R., TROAKES, C., KING, A., SAHNI, V., DE JONG, S., BOSSERS, K., PAPOULI, E., MIRZA, M., AL-SARRAJ, S., SHAW, C. E., SHAW, P. J., KIRBY, J., VELDINK, J. H., MACKLIS, J. D., POWELL, J. F. & AL-CHALABI, A. 2015. Stratified gene expression analysis identifies major amyotrophic lateral sclerosis genes. *Neurobiol Aging,* 36**,** 2006 e1-9.

JONES, J. I. & CLEMMONS, D. R. 1995. Insulin-like growth factors and their binding proteins: biological actions. *Endocr Rev,* 16**,** 3-34.

JOSS-MOORE, L. A. & LANE, R. H. 2012. Epigenetics and the developmental origins of disease: the key to unlocking the door of personalized medicine. *Epigenomics,* 4**,** 471-3.

JOVANOVIC, L. 2001. Continuous glucose monitoring during pregnancy complicated by gestational diabetes mellitus. *Curr Diab Rep,* 1**,** 82-5.

JOVANOVIC, L. & PETTITT, D. J. 2001. Gestational diabetes mellitus. *JAMA,* 286**,** 2516-8.

JUUL, A. 2003. Serum levels of insulin-like growth factor I and its binding proteins in health and disease. *Growth Horm IGF Res,* 13**,** 113-70.

KAAJA, R. & RONNEMAA, T. 2008. Gestational diabetes: pathogenesis and consequences to mother and offspring. *Rev Diabet Stud,* 5**,** 194-202.

KABIR, G. H., M; FARAQUE, O; ALAUDDIN, M AND ALI, L. 2014. Association of serum free IGF-1 and IGFBP-1 with insulin sensitivity and insulin secretory defects in Bangladeshi type 2 diabetes mellitus *Journal of Taibah University Medical Sciences,* 9**,** 132-138.

KALRA, S., GUPTA, Y. & KUMAR, A. 2016. Prevention of Gestational Diabetes Mellitus (GDM). *J Pak Med Assoc,* 66**,** S107-9.

KALTER-LEIBOVICI, O., FREEDMAN, L. S., OLMER, L., LIEBERMANN, N., HEYMANN, A., TAL, O., LERNER-GEVA, L., MELAMED, N. & HOD, M. 2012. Screening and diagnosis of gestational diabetes mellitus: critical appraisal of the new International Association of Diabetes in Pregnancy Study Group recommendations on a national level. *Diabetes Care,* 35**,** 1894-6.

KAMMEL, A., SAUSSENTHALER, S., JAHNERT, M., JONAS, W., STIRM, L., HOEFLICH, A., STAIGER, H., FRITSCHE, A., HARING, H. U., JOOST, H. G., SCHURMANN, A. & SCHWENK, R. W. 2016. Early hypermethylation of hepatic Igfbp2 results in its reduced expression preceding fatty liver in mice. *Hum Mol Genet,* 25**,** 2588-2599.

KANHERKAR, R. R., BHATIA-DEY, N. & CSOKA, A. B. 2014. Epigenetics across the human lifespan. *Front Cell Dev Biol,* 2**,** 49.

KARAGIANNIS, T., BEKIARI, E., MANOLOPOULOS, K., PALETAS, K. & TSAPAS, A. 2010. Gestational diabetes mellitus: why screen and how to diagnose. *Hippokratia,* 14**,** 151-4.

KARATZAS, P. S., GAZOULI, M., SAFIOLEAS, M. & MANTZARIS, G. J. 2014. DNA methylation changes in inflammatory bowel disease. *Ann Gastroenterol,* 27**,** 125-132.

KASUGA, M. 2011. KCNQ1, a susceptibility gene for type 2 diabetes. *J Diabetes Investig,* 2**,** 413-4.

KHOSRAVI, J., KRISHNA, R. G., BODANI, U., DIAMANDI, A., KHAJA, N., KALRA, B. & KUMAR, A. 2007. Immunoassay of serine-phosphorylated isoform of insulin-like growth factor (IGF) binding protein (IGFBP)-1. *Clin Biochem,* 40**,** 86-93.

KHULAN, B., COOPER, W. N., SKINNER, B. M., BAUER, J., OWENS, S., PRENTICE, A. M., BELTEKI, G., CONSTANCIA, M., DUNGER, D. & AFFARA, N. A. 2012. Periconceptional maternal micronutrient supplementation is associated with widespread gender related changes in the epigenome: a study of a unique resource in the Gambia. *Hum Mol Genet,* 21**,** 2086-101.

KIM, S. Y., SHARMA, A. J. & CALLAGHAN, W. M. 2012. Gestational diabetes and childhood obesity: what is the link? *Curr Opin Obstet Gynecol,* 24**,** 376-81.

KOUKOURA, O., SIFAKIS, S. & SPANDIDOS, D. A. 2012. DNA methylation in the human placenta and fetal growth (review). *Mol Med Rep,* 5**,** 883-9.

KOUTSAKI, M., SIFAKIS, S., ZARAVINOS, A., KOUTROULAKIS, D., KOUKOURA, O. & SPANDIDOS, D. A. 2011. Decreased placental expression of hPGH, IGF-I and IGFBP-1 in pregnancies complicated by fetal growth restriction. *Growth Horm IGF Res,* 21**,** 31-6.

KRATZ, A. & CARNINCI, P. 2014. The devil in the details of RNA-seq. *Nat Biotechnol,* 32**,** 882-4.

KRJUTSKOV, K., KOEL, M., ROOST, A. M., KATAYAMA, S., EINARSDOTTIR, E., JOUHILAHTI, E. M., SODERHALL, C., JAAKMA, U., PLAAS, M., VESTERLUND, L., LOHI, H., SALUMETS, A. & KERE, J. 2016. Globin mRNA reduction for whole-blood transcriptome sequencing. *Sci Rep,* 6**,** 31584.

LABAJ, P. P., LEPARC, G. G., LINGGI, B. E., MARKILLIE, L. M., WILEY, H. S. & KREIL, D. P. 2011. Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics,* 27**,** i383-91.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LAPPAS, M. 2015. Insulin-like growth factor-binding protein 1 and 7 concentrations are lower in obese pregnant women, women with gestational diabetes and their fetuses. *J Perinatol,* 35**,** 32-8.

LAPPAS, M., HIDEN, U., DESOYE, G., FROEHLICH, J., HAUGUEL-DE MOUZON, S. & JAWERBAUM, A. 2011. The role of oxidative stress in the pathophysiology of gestational diabetes mellitus. *Antioxid Redox Signal,* 15**,** 3061-100.

LASKA, M. J., NISSEN, K. K. & NEXO, B. A. 2013. (Some) cellular mechanisms influencing the transcription of human endogenous retrovirus, HERV-Fc1. *PLoS One,* 8**,** e53895.

LAW, J. A. & JACOBSEN, S. E. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet,* 11**,** 204-20.

LAW, J. R., STAFFORD, J. M., D'AGOSTINO, R. B., JR., BADARU, A., CRUME, T. L., DABELEA, D., DOLAN, L. M., LAWRENCE, J. M., PETTITT, D. J. & MAYER-DAVIS, E. J. 2015. Association of parental history of diabetes with cardiovascular disease risk factors in children with type 2 diabetes. *J Diabetes Complications,* 29**,** 534-9.

LEE, P. D., LUSTIG, R. H., LENDERS, C., BAILLARGEON, J., WILSON, D. M. & GLASER PEDIATRIC RESEARCH NETWORK OBESITY STUDY, G. 2016. Insulin-Like Growth Factor Binding Protein 1 Predicts Insulin Sensitivity and Insulin Area-under-the-Curve in Obese, Nondiabetic Adolescents. *Endocr Pract,* 22**,** 136-42.

LEE, S. K., SON LE, T., CHOI, H. J. & AHNN, J. 2013. Dicarbonyl/l-xylulose reductase (DCXR): The multifunctional pentosuria enzyme. *Int J Biochem Cell Biol,* 45**,** 2563-7.

LEHNEN, H., ZECHNER, U. & HAAF, T. 2013. Epigenetics of gestational diabetes mellitus and offspring health: the time for action is in early stages of life. *Mol Hum Reprod,* 19**,** 415-22.

LIAO, S., LIU, Y., TAN, Y., GAN, L., MEI, J., SONG, W., CHI, S., DONG, X., CHEN, X. & DENG, S. 2012. Association of genetic variants of melatonin receptor 1B with gestational plasma glucose level and risk of glucose intolerance in pregnant Chinese women. *PLoS One,* 7**,** e40113.

LIGUORI, A., PUGLIANIELLO, A., GERMANI, D., DEODATI, A., PESCHIAROLI, E. & CIANFARANI, S. 2010. Epigenetic changes predisposing to type 2 diabetes in intrauterine growth retardation. *Front Endocrinol (Lausanne),* 1**,** 5.

LINDSAY, R. S., WESTGATE, J. A., BEATTIE, J., PATTISON, N. S., GAMBLE, G., MILDENHALL, L. F., BREIER, B. H. & JOHNSTONE, F. D. 2007. Inverse changes in fetal insulin-like growth factor (IGF)-1 and IGF binding protein-1 in association with higher birth weight in maternal diabetes. *Clin Endocrinol (Oxf),* 66**,** 322-8.

LIU, J., WALTER, E., STENGER, D. & THACH, D. 2006. Effects of globin mRNA reduction methods on gene expression profiles from whole blood. *J Mol Diagn,* 8**,** 551-8.

LIU, Y., MORLEY, M., BRANDIMARTO, J., HANNENHALLI, S., HU, Y., ASHLEY, E. A., TANG, W. H., MORAVEC, C. S., MARGULIES, K. B., CAPPOLA, T. P., LI, M. & CONSORTIUM, M. A. 2015. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics,* 105**,** 83-9.

LIU, Y. J., TSUSHIMA, T., ONODA, N., MINEI, S., SANAKA, M., NAGASHIMA, T., YANAGISAWA, K. & OMORI, Y. 1996. Expression of messenger RNA of insulin-like growth factors (IGFs) and IGF binding proteins (IGFBP1-6) in placenta of normal and diabetic pregnancy. *Endocr J,* 43 Suppl**,** S89-91.

LOUKOVAARA, S., KAAJA, R. J. & KOISTINEN, R. A. 2005. Cord serum insulin-like growth factor binding protein-1 and -3: effect of maternal diabetes and relationships to fetal growth. *Diabetes Metab,* 31**,** 163-7.

LOWE, L. P., METZGER, B. E., DYER, A. R., LOWE, J., MCCANCE, D. R., LAPPIN, T. R., TRIMBLE, E. R., COUSTAN, D. R., HADDEN, D. R., HOD, M., OATS, J. J., PERSSON, B. & GROUP, H. S. C. R. 2012. Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: associations of maternal A1C and glucose with pregnancy outcomes. *Diabetes Care,* 35**,** 574-80.

LOWE, R. & RAKYAN, V. K. 2014. Correcting for cell-type composition bias in epigenome-wide association studies. *Genome Med,* 6**,** 23.

LU, S., PUROHIT, S., SHARMA, A., ZHI, W., HE, M., WANG, Y., LI, C. J. & SHE, J. X. 2012. Serum insulin-like growth factor binding protein 6 (IGFBP6) is increased in patients with type 1 diabetes and its complications. *Int J Clin Exp Med,* 5**,** 229-37.

LUMEY, L. H., STEIN, A. D., KAHN, H. S., VAN DER PAL-DE BRUIN, K. M., BLAUW, G. J., ZYBERT, P. A. & SUSSER, E. S. 2007. Cohort profile: the Dutch Hunger Winter families study. *Int J Epidemiol,* 36**,** 1196-204.

MACAULAY, S., DUNGER, D. B. & NORRIS, S. A. 2014. Gestational diabetes mellitus in Africa: a systematic review. *PLoS One,* 9**,** e97871.

MATOUGH, F. A., BUDIN, S. B., HAMID, Z. A., ALWAHAIBI, N. & MOHAMED, J. 2012. The role of oxidative stress and antioxidants in diabetic complications. *Sultan Qaboos Univ Med J,* 12**,** 5-18.

MBANYA, J. C., SQUIRE, S. B., CAZAP, E. & PUSKA, P. 2011. Mobilising the world for chronic NCDs. *Lancet,* 377**,** 536-7.

MCCARTHY, N. S., MELTON, P. E., CADBY, G., YAZAR, S., FRANCHINA, M., MOSES, E. K., MACKEY, D. A. & HEWITT, A. W. 2014. Meta-analysis of human methylation data for evidence of sex-specific autosomal patterns. *BMC Genomics,* 15**,** 981.

MCGUINNESS, D., MCGLYNN, L. M., JOHNSON, P. C., MACINTYRE, A., BATTY, G. D., BURNS, H., CAVANAGH, J., DEANS, K. A., FORD, I., MCCONNACHIE, A., MCGINTY, A., MCLEAN, J. S., MILLAR, K., PACKARD, C. J., SATTAR, N. A., TANNAHILL, C., VELUPILLAI, Y. N. & SHIELS, P. G. 2012. Socio-economic status is associated with epigenetic differences in the pSoBid cohort. *Int J Epidemiol,* 41**,** 151-60.

MCRAE, A. F., POWELL, J. E., HENDERS, A. K., BOWDLER, L., HEMANI, G., SHAH, S., PAINTER, J. N., MARTIN, N. G., VISSCHER, P. M. & MONTGOMERY, G. W. 2014. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol,* 15**,** R73.

MEDRIKOVA, D., SIJMONSMA, T. P., SOWODNIOK, K., RICHARDS, D. M., DELACHER, M., STICHT, C., GRETZ, N., SCHAFMEIER, T., FEUERER, M. & HERZIG, S. 2015. Brown adipose tissue harbors a distinct sub-population of regulatory T cells. *PLoS One,* 10**,** e0118534.

MEGIA, A., VENDRELL, J., GUTIERREZ, C., SABATE, M., BROCH, M., FERNANDEZ-REAL, J. M. & SIMON, I. 2008. Insulin sensitivity and resistin levels in gestational diabetes mellitus and after parturition. *Eur J Endocrinol,* 158**,** 173-8.

MESSERSCHMIDT, D. M., KNOWLES, B. B. & SOLTER, D. 2014. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev,* 28**,** 812-28.

METZGER, B. E., LOWE, L. P., DYER, A. R., TRIMBLE, E. R., CHAOVARINDR, U., COUSTAN, D. R., HADDEN, D. R., MCCANCE, D. R., HOD, M., MCINTYRE, H. D., OATS, J. J., PERSSON, B., ROGERS, M. S. & SACKS, D. A. 2008. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med,* 358**,** 1991-2002.

MEUGNIER, E., FARAJ, M., ROME, S., BEAUREGARD, G., MICHAUT, A., PELLOUX, V., CHIASSON, J. L., LAVILLE, M., CLEMENT, K., VIDAL, H. & RABASA-LHORET, R. 2007. Acute hyperglycemia induces a global downregulation of gene expression in adipose tissue and skeletal muscle of healthy subjects. *Diabetes,* 56**,** 992-9.

MICHALEK, R. D., GERRIETS, V. A., JACOBS, S. R., MACINTYRE, A. N., MACIVER, N. J., MASON, E. F., SULLIVAN, S. A., NICHOLS, A. G. & RATHMELL, J. C. 2011. Cutting edge: distinct glycolytic and lipid oxidative metabolic programs are essential for effector and regulatory CD4+ T cell subsets. *J Immunol,* 186**,** 3299-303.

MILLER, C. L., DIGLISIC, S., LEISTER, F., WEBSTER, M. & YOLKEN, R. H. 2004. Evaluating RNA status for RT-PCR in extracts of postmortem human brain tissue. *Biotechniques,* 36**,** 628-33.

MILLER, S. A., DYKES, D. D. & POLESKY, H. F. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res,* 16**,** 1215.

MOORE, L. D., LE, T. & FAN, G. 2013. DNA methylation and its basic function. *Neuropsychopharmacology,* 38**,** 23-38.

MORITA, S., HORII, T., KIMURA, M. & HATADA, I. 2013. MiR-184 regulates insulin secretion through repression of Slc25a22. *PeerJ,* 1**,** e162.
MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* 5**,** 621-8.
MOSSIE, T. B., BERHE, G. H., KAHSAY, G. H. & TAREKE, M. 2017. Prevalence of Depression and Associated Factors among Diabetic Patients at Mekelle City, North Ethiopia. *Indian J Psychol Med,* 39**,** 52-58.
MUFUNDA, J., CHATORA, R., NDAMBAKUWA, Y., NYARANGO, P., KOSIA, A., CHIFAMBA, J., FILIPE, A., USMAN, A. & SPARKS, V. H. 2006. Emerging non-communicable disease epidemic in Africa: preventive measures from the WHO Regional Office for Africa. *Ethn Dis,* 16**,** 521-6.
MUREA, M., MA, L. & FREEDMAN, B. I. 2012. Genetic and environmental factors associated with type 2 diabetes and diabetic vascular complications. *Rev Diabet Stud,* 9**,** 6-22.
MUTRYN, M. F., BRANNICK, E. M., FU, W., LEE, W. R. & ABASHT, B. 2015. Characterization of a novel chicken muscle disorder through differential gene expression and pathway analysis using RNA-sequencing. *BMC Genomics,* 16**,** 399.
MWANRI, A. W., KINABO, J., RAMAIYA, K. & FESKENS, E. J. 2015. Gestational diabetes mellitus in sub-Saharan Africa: systematic review and metaregression on prevalence and risk factors. *Trop Med Int Health,* 20**,** 983-1002.
MYATT, L. & CUI, X. 2004. Oxidative stress in the placenta. *Histochem Cell Biol,* 122**,** 369-82.
NACHUM, Z., BEN-SHLOMO, I., WEINER, E. & SHALEV, E. 1999. Twice daily versus four times daily insulin dose regimens for diabetes in pregnancy: randomised controlled trial. *BMJ,* 319**,** 1223-7.
NACU, S., YUAN, W., KAN, Z., BHATT, D., RIVERS, C. S., STINSON, J., PETERS, B. A., MODRUSAN, Z., JUNG, K., SESHAGIRI, S. & WU, T. D. 2011. Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. *BMC Med Genomics,* 4**,** 11.
NAWATHE, A. & LEES, C. 2016. Early onset fetal growth restriction. *Best Pract Res Clin Obstet Gynaecol*.
NAWATHE, A. R., CHRISTIAN, M., KIM, S. H., JOHNSON, M., SAVVIDOU, M. D. & TERZIDOU, V. 2016. Insulin-like growth factor axis in pregnancies affected by fetal growth disorders. *Clin Epigenetics,* 8**,** 11.
NIEMINEN, T. T., SHOMAN, S., EISSA, S., PELTOMAKI, P. & ABDEL-RAHMAN, W. M. 2012. Distinct genetic and epigenetic signatures of colorectal cancers according to ethnic origin. *Cancer Epidemiol Biomarkers Prev,* 21**,** 202-11.
NIIRANEN, O. 2015. *RNA-sequencing reveals stromal cell induced changes in cancer cell gene expression profiles*

Bachelor of Engineering, Helsinki Metropolia University of Applied Sciences

NOCTOR, E. & DUNNE, F. P. 2015. Type 2 diabetes after gestational diabetes: The influence of changing diagnostic criteria. *World J Diabetes,* 6**,** 234-44.
O'SULLIVAN, J. B. & MAHAN, C. M. 1964. Criteria for the Oral Glucose Tolerance Test in Pregnancy. *Diabetes,* 13**,** 278-85.
OATS, J. N. & BEISCHER, N. A. 1986. Gestational diabetes. *Aust N Z J Obstet Gynaecol,* 26**,** 2-10.
OSEI-KUMAH, A., SMITH, R., JURISICA, I., CANIGGIA, I. & CLIFTON, V. L. 2011. Sex-specific differences in placental global gene expression in pregnancies complicated by asthma. *Placenta,* 32**,** 570-8.

OSEI, K., SCHUSTER, D. P., AMOAH, A. G. & OWUSU, S. K. 2003. Diabetes in Africa. Pathogenesis of type 1 and type 2 diabetes mellitus in sub-Saharan Africa: implications for transitional populations. *J Cardiovasc Risk,* 10**,** 85-96.

OSHLACK, A., ROBINSON, M. D. & YOUNG, M. D. 2010. From RNA-seq reads to differential expression results. *Genome Biol,* 11**,** 220.

OUATTARA, A. K., BISSEYE, C., BAZIE, B. V., DIARRA, B., COMPAORE, T. R., DJIGMA, F., PIETRA, V., MORET, R. & SIMPORE, J. 2014. Glucose-6-phosphate dehydrogenase (G6PD) deficiency is associated with asymptomatic malaria in a rural community in Burkina Faso. *Asian Pac J Trop Biomed,* 4**,** 655-8.

OZSOLAK, F. & MILOS, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet,* 12**,** 87-98.

PACAL, L., TOMANDL, J., SVOJANOVSKY, J., KRUSOVA, D., STEPANKOVA, S., REHOROVA, J., OLSOVSKY, J., BELOBRADKOVA, J., TANHAUSEROVA, V., TOMANDLOVA, M., MUZIK, J. & KANKOVA, K. 2011. Role of thiamine status and genetic variability in transketolase and other pentose phosphate cycle enzymes in the progression of diabetic nephropathy. *Nephrol Dial Transplant,* 26**,** 1229-36.

PARK, J. H., STOFFERS, D. A., NICHOLLS, R. D. & SIMMONS, R. A. 2008. Development of type 2 diabetes following intrauterine growth retardation in rats is associated with progressive epigenetic silencing of Pdx1. *J Clin Invest,* 118**,** 2316-24.

PATRA, K. C. & HAY, N. 2014. The pentose phosphate pathway and cancer. *Trends Biochem Sci,* 39**,** 347-54.

PAUL, D. S. & BECK, S. 2014. Advances in epigenome-wide association studies for common diseases. *Trends Mol Med*.

PEARCE, M. S., MANN, K. D., RELTON, C. L., FRANCIS, R. M., STEELE, J. G., CRAFT, A. W. & PARKER, L. 2012. How the Newcastle Thousand Families birth cohort study has contributed to the understanding of the impact of birth weight and early life socioeconomic position on disease in later life. *Maturitas,* 72**,** 23-8.

PERKINS, J. R., DAWES, J. M., MCMAHON, S. B., BENNETT, D. L., ORENGO, C. & KOHL, M. 2012. ReadqPCR and NormqPCR: R packages for the reading, quality checking and normalisation of RT-qPCR quantification cycle (Cq) data. *BMC Genomics,* 13**,** 296.

PERNER, A., NIELSEN, S. E. & RASK-MADSEN, J. 2003. High glucose impairs superoxide production from isolated blood neutrophils. *Intensive Care Med,* 29**,** 642-5.

PETERS, A. L. & VAN NOORDEN, C. J. 2009. Glucose-6-phosphate dehydrogenase deficiency and malaria: cytochemical detection of heterozygous G6PD deficiency in women. *J Histochem Cytochem,* 57**,** 1003-11.

PETRONIS, A. 2010. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature,* 465**,** 721-7.

PETRY, C. J. 2010. Gestational diabetes: risk factors and recent advances in its genetics and treatment. *Br J Nutr,* 104**,** 775-87.

PETTITT, D. J., LAWRENCE, J. M., BEYER, J., HILLIER, T. A., LIESE, A. D., MAYER-DAVIS, B., LOOTS, B., IMPERATORE, G., LIU, L., DOLAN, L. M., LINDER, B. & DABELEA, D. 2008. Association between maternal diabetes in utero and age at offspring's diagnosis of type 2 diabetes. *Diabetes Care,* 31**,** 2126-30.

PINNA, A., CONTINI, E. L., CARRU, C. & SOLINAS, G. 2013. Glucose-6-phosphate dehydrogenase deficiency and diabetes mellitus with severe retinal complications in a Sardinian population, Italy. *Int J Med Sci,* 10**,** 1907-13.

PINNEY, S. E. & SIMMONS, R. A. 2012. Metabolic programming, epigenetics, and gestational diabetes mellitus. *Curr Diab Rep,* 12**,** 67-74.

PLAGEMANN, A., HARDER, T., BRUNN, M., HARDER, A., ROEPKE, K., WITTROCK-STAAR, M., ZISKA, T., SCHELLONG, K., RODEKAMP, E., MELCHIOR, K. & DUDENHAUSEN, J. W. 2009. Hypothalamic proopiomelanocortin promoter methylation becomes altered by early overfeeding: an epigenetic model of obesity and the metabolic syndrome. *J Physiol,* 587**,** 4963-76.

PON, C. K., FIRTH, S. M. & BAXTER, R. C. 2015. Involvement of insulin-like growth factor binding protein-3 in peroxisome proliferator-activated receptor gamma-mediated inhibition of breast cancer cell growth. *Mol Cell Endocrinol,* 399**,** 354-61.

POULAKOS, P., MINTZIORI, G., TSIROU, E., TAOUSANI, E., SAVVAKI, D., HARIZOPOULOU, V. & GOULIS, D. G. 2015. Comments on gestational diabetes mellitus: from pathophysiology to clinical practice. *Hormones (Athens),* 14**,** 335-44.

PUOANE, T. T., L; SANDERS, D; PARKER, W 2008. Chronic non-communicable diseases. *In:* IN: BARRON, P. A. R.-R., J, (EDS). (ed.) *South African Health Review 2008. Durban: Health Systems Trust.*

QIU, H. Y., CHU, Y. L., LI, M., SUN, Y. Y. & LI, H. F. 2005. [Tyrosine phosphorylation and protein expression of insulin receptor substrate-2 in the adipose tissue from patients with polycystic ovary syndrome]. *Zhonghua Fu Chan Ke Za Zhi,* 40**,** 116-9.

RAHAL, A., KUMAR, A., SINGH, V., YADAV, B., TIWARI, R., CHAKRABORTY, S. & DHAMA, K. 2014. Oxidative stress, prooxidants, and antioxidants: the interplay. *Biomed Res Int,* 2014**,** 761264.

RAJPATHAK, S. N., GUNTER, M. J., WYLIE-ROSETT, J., HO, G. Y., KAPLAN, R. C., MUZUMDAR, R., ROHAN, T. E. & STRICKLER, H. D. 2009. The role of insulin-like growth factor-I and its binding proteins in glucose homeostasis and type 2 diabetes. *Diabetes Metab Res Rev,* 25**,** 3-12.

RAJWANI, A., EZZAT, V., SMITH, J., YULDASHEVA, N. Y., DUNCAN, E. R., GAGE, M., CUBBON, R. M., KAHN, M. B., IMRIE, H., ABBAS, A., VISWAMBHARAN, H., AZIZ, A., SUKUMAR, P., VIDAL-PUIG, A., SETHI, J. K., XUAN, S., SHAH, A. M., GRANT, P. J., PORTER, K. E., KEARNEY, M. T. & WHEATCROFT, S. B. 2012. Increasing circulating IGFBP1 levels improves insulin sensitivity, promotes nitric oxide production, lowers blood pressure, and protects against atherosclerosis. *Diabetes,* 61**,** 915-24.

RAKYAN, V. K., DOWN, T. A., BALDING, D. J. & BECK, S. 2011. Epigenome-wide association studies for common human diseases. *Nat Rev Genet,* 12**,** 529-41.

RANI, P. R. & BEGUM, J. 2016. Screening and Diagnosis of Gestational Diabetes Mellitus, Where Do We Stand. *J Clin Diagn Res,* 10**,** QE01-4.

RAPAPORT, F., KHANIN, R., LIANG, Y., PIRUN, M., KREK, A., ZUMBO, P., MASON, C. E., SOCCI, N. D. & BETEL, D. 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol,* 14**,** R95.

RELTON, C. L., GROOM, A., ST POURCAIN, B., SAYERS, A. E., SWAN, D. C., EMBLETON, N. D., PEARCE, M. S., RING, S. M., NORTHSTONE, K., TOBIAS, J. H., TRAKALO, J., NESS, A. R., SHAHEEN, S. O. & DAVEY SMITH, G. 2012. DNA methylation patterns in cord blood DNA and body size in childhood. *PLoS One,* 7**,** e31821.

RELTON, C & DAVEY SMITH, G. 2012. Two-step epigentic mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 41(4) 161-176

RETNAKARAN, R. 2016. The Insulin-Like Growth Factor Axis: A New Player in Gestational Diabetes Mellitus? *Diabetes,* 65**,** 3246-3248.

RINAUDO, P. & WANG, E. 2012. Fetal programming and metabolic syndrome. *Annu Rev Physiol,* 74**,** 107-30.

ROBITAILLE, J. & GRANT, A. M. 2008. The genetics of gestational diabetes mellitus: evidence for relationship with type 2 diabetes mellitus. *Genet Med,* 10**,** 240-50.

ROGLIC, G., UNWIN, N., BENNETT, P. H., MATHERS, C., TUOMILEHTO, J., NAG, S., CONNOLLY, V. & KING, H. 2005. The burden of mortality attributable to diabetes: realistic estimates for the year 2000. *Diabetes Care,* 28**,** 2130-5.

ROY, S. W. & IRIMIA, M. 2008. When good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis. *Bioessays,* 30**,** 601-5.

RUAN, W. & LAI, M. 2010. Insulin-like growth factor binding protein: a possible marker for the metabolic syndrome? *Acta Diabetol,* 47**,** 5-14.

RUCHAT, S. M., HIVERT, M. F. & BOUCHARD, L. 2013a. Epigenetic programming of obesity and diabetes by in utero exposure to gestational diabetes mellitus. *Nutr Rev,* 71 Suppl 1**,** S88-94.

RUCHAT, S. M., HOUDE, A. A., VOISIN, G., ST-PIERRE, J., PERRON, P., BAILLARGEON, J. P., GAUDET, D., HIVERT, M. F., BRISSON, D. & BOUCHARD, L. 2013b. Gestational diabetes mellitus epigenetically affects genes predominantly involved in metabolic diseases. *Epigenetics,* 8**,** 935-43.

SADAKIERSKA-CHUDY, A. & FILIP, M. 2015. A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs. *Neurotox Res,* 27**,** 172-97.

SADIKOVIC, B., AL-ROMAIH, K., SQUIRE, J. A. & ZIELENSKA, M. 2008. Cause and consequences of genetic and epigenetic alterations in human cancer. *Curr Genomics,* 9**,** 394-408.

SAEED, T. K., HAMAMY, H. A. & ALWAN, A. A. 1985. Association of glucose-6-phosphate dehydrogenase deficiency with diabetes mellitus. *Diabet Med,* 2**,** 110-2.

SAMUELSSON, A. M., MATTHEWS, P. A., ARGENTON, M., CHRISTIE, M. R., MCCONNELL, J. M., JANSEN, E. H., PIERSMA, A. H., OZANNE, S. E., TWINN, D. F., REMACLE, C., ROWLERSON, A., POSTON, L. & TAYLOR, P. D. 2008. Diet-induced obesity in female mice leads to offspring hyperphagia, adiposity, hypertension, and insulin resistance: a novel murine model of developmental programming. *Hypertension,* 51**,** 383-92.

SAMUELSSON, A. M., MATTHEWS, P. A., JANSEN, E., TAYLOR, P. D. & POSTON, L. 2013. Sucrose feeding in mouse pregnancy leads to hypertension, and sex-linked obesity and insulin resistance in female offspring. *Front Physiol,* 4**,** 14.

SANABRIA-MARTINEZ, G., GARCIA-HERMOSO, A., POYATOS-LEON, R., ALVAREZ-BUENO, C., SANCHEZ-LOPEZ, M. & MARTINEZ-VIZCAINO, V. 2015. Effectiveness of physical activity interventions on preventing gestational diabetes mellitus and excessive maternal weight gain: a meta-analysis. *BJOG,* 122**,** 1167-74.

SANDHU, M. S., DUNGER, D. B. & GIOVANNUCCI, E. L. 2002. Insulin, insulin-like growth factor-I (IGF-I), IGF binding proteins, their biologic interactions, and colorectal cancer. *J Natl Cancer Inst,* 94**,** 972-80.

SCHNEIDER, A. M., RAWAT, D., WEINSTEIN, L. S., GUPTE, S. A. & RICHARDS, W. O. 2012. Effects of laparoscopic Roux-en-Y gastric bypass on glucose-6 phosphate dehydrogenase activity in obese type 2 diabetics. *Surg Endosc,* 26**,** 823-30.

SCHNEIDER, H., SHAW, J. & ZIMMET, P. 2003. Guidelines for the detection of diabetes mellitus--diagnostic criteria and rationale for screening. *Clin Biochem Rev,* 24**,** 77-80.

SENDLER, E., JOHNSON, G. D. & KRAWETZ, S. A. 2011. Local and global factors affecting RNA sequencing analysis. *Anal Biochem,* 419**,** 317-22.

SERMAN, L., VLAHOVIC, M., SIJAN, M., BULIC-JAKUS, F., SERMAN, A., SINCIC, N., MATIJEVIC, R., JURIC-LEKIC, G. & KATUSIC, A. 2007. The impact of 5-azacytidine on placental weight, glycoprotein pattern and proliferating cell nuclear antigen expression in rat placenta. *Placenta,* 28**,** 803-11.

SFERRUZZI-PERRI, A. N., OWENS, J. A., PRINGLE, K. G. & ROBERTS, C. T. 2011. The neglected role of insulin-like growth factors in the maternal circulation regulating fetal growth. *J Physiol,* 589**,** 7-20.

SHAAT, N. & GROOP, L. 2007. Genetics of gestational diabetes mellitus. *Curr Med Chem,* 14**,** 569-83.
SHAM, P. C. & PURCELL, S. M. 2014. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet,* 15**,** 335-46.
SHAW, J. E., SICREE, R. A. & ZIMMET, P. Z. 2010. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract,* 87**,** 4-14.
SHEN, X., XI, G., WAI, C. & CLEMMONS, D. R. 2015. The coordinate cellular response to insulin-like growth factor-I (IGF-I) and insulin-like growth factor-binding protein-2 (IGFBP-2) is regulated through vimentin binding to receptor tyrosine phosphatase beta (RPTPbeta). *J Biol Chem,* 290**,** 11578-90.
SHENDURE, J. 2008. The beginning of the end for microarrays? *Nat Methods,* 5**,** 585-7.
SHENDURE, J. & JI, H. 2008. Next-generation DNA sequencing. *Nat Biotechnol,* 26**,** 1135-45.
SHIN, H., SHANNON, C. P., FISHBANE, N., RUAN, J., ZHOU, M., BALSHAW, R., WILSON-MCMANUS, J. E., NG, R. T., MCMANUS, B. M., TEBBUTT, S. J. & TEAM, P. C. O. E. 2014. Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PLoS One,* 9**,** e91041.
SOUMYA, S., ROHILLA, M., CHOPRA, S., DUTTA, S., BHANSALI, A., PARTHAN, G. & DUTTA, P. 2015. HbA1c: A Useful Screening Test for Gestational Diabetes Mellitus. *Diabetes Technol Ther,* 17**,** 899-904.
SPEGEL, P., SHAROYKO, V. V., GOEHRING, I., DANIELSSON, A. P., MALMGREN, S., NAGORNY, C. L., ANDERSSON, L. E., KOECK, T., SHARP, G. W., STRAUB, S. G., WOLLHEIM, C. B. & MULDER, H. 2013. Time-resolved metabolomics analysis of beta-cells implicates the pentose phosphate pathway in the control of insulin release. *Biochem J,* 450**,** 595-605.
SRINIVASAN, S., BOLICK, D. T., HATLEY, M. E., NATARAJAN, R., REILLY, K. B., YEH, M., CHRESTENSEN, C., STURGILL, T. W. & HEDRICK, C. C. 2004. Glucose regulates interleukin-8 production in aortic endothelial cells through activation of the p38 mitogen-activated protein kinase pathway in diabetes. *J Biol Chem,* 279**,** 31930-6.
STANTON, R. C. 2012. Glucose-6-phosphate dehydrogenase, NADPH, and cell survival. *IUBMB Life,* 64**,** 362-9.
STINCONE, A., PRIGIONE, A., CRAMER, T., WAMELINK, M. M., CAMPBELL, K., CHEUNG, E., OLIN-SANDOVAL, V., GRUNING, N. M., KRUGER, A., TAUQEER ALAM, M., KELLER, M. A., BREITENBACH, M., BRINDLE, K. M., RABINOWITZ, J. D. & RALSER, M. 2015. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biol Rev Camb Philos Soc,* 90**,** 927-63.
STIRZAKER, C., TABERLAY, P. C., STATHAM, A. L. & CLARK, S. J. 2014. Mining cancer methylomes: prospects and challenges. *Trends Genet,* 30**,** 75-84.
SULTAN, M., DOKEL, S., AMSTISLAVSKIY, V., WUTTIG, D., SULTMANN, H., LEHRACH, H. & YASPO, M. L. 2012. A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. *Biochem Biophys Res Commun,* 422**,** 643-6.
SUN, X. Q., LUO, Y. Y., AN, L. W., CHU, L., HUO, L. L., HAN, X. Y., ZHOU, X. H., REN, Q. & JI, L. N. 2011. Contribution of the Akt2 gene to type 2 diabetes in the Chinese Han population. *Chin Med J (Engl),* 124**,** 725-8.
TANHAUSEROVA, V., KURICOVA, K., PACAL, L., BARTAKOVA, V., REHOROVA, J., SVOJANOVSKY, J., OLSOVSKY, J., BELOBRADKOVA, J. & KANKOVA, K. 2014. Genetic variability in enzymes of metabolic pathways conferring protection against non-enzymatic glycation versus diabetes-related morbidity and mortality. *Clin Chem Lab Med,* 52**,** 77-83.
TATE, P. H. & BIRD, A. P. 1993. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev,* 3**,** 226-31.

TEH, A. L., PAN, H., CHEN, L., ONG, M. L., DOGRA, S., WONG, J., MACISAAC, J. L., MAH, S. M., MCEWEN, L. M., SAW, S. M., GODFREY, K. M., CHONG, Y. S., KWEK, K., KWOH, C. K., SOH, S. E., CHONG, M. F., BARTON, S., KARNANI, N., CHEONG, C. Y., BUSCHDORF, J. P., STUNKEL, W., KOBOR, M. S., MEANEY, M. J., GLUCKMAN, P. D. & HOLBROOK, J. D. 2014. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res,* 24**,** 1064-74.

TEKOLA-AYELE, F., ADEYEMO, A. A. & ROTIMI, C. N. 2013. Genetic epidemiology of type 2 diabetes and cardiovascular diseases in Africa. *Prog Cardiovasc Dis,* 56**,** 251-60.

TIAN, W. N., BRAUNSTEIN, L. D., PANG, J., STUHLMEIER, K. M., XI, Q. C., TIAN, X. & STANTON, R. C. 1998. Importance of glucose-6-phosphate dehydrogenase activity for cell growth. *J Biol Chem,* 273**,** 10609-17.

TOBI, E. W., LUMEY, L. H., TALENS, R. P., KREMER, D., PUTTER, H., STEIN, A. D., SLAGBOOM, P. E. & HEIJMANS, B. T. 2009. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet,* 18**,** 4046-53.

TOBIAS, D. K., ZHANG, C., VAN DAM, R. M., BOWERS, K. & HU, F. B. 2011. Physical activity before and during pregnancy and risk of gestational diabetes mellitus: a meta-analysis. *Diabetes Care,* 34**,** 223-9.

TOBIAS, M. 2011. Global control of diabetes: information for action. *Lancet,* 378**,** 3-4.

TRAPNELL, C., HENDRICKSON, D. G., SAUVAGEAU, M., GOFF, L., RINN, J. L. & PACHTER, L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol,* 31**,** 46-53.

TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc,* 7**,** 562-78.

TREROTOLA, M., RELLI, V., SIMEONE, P. & ALBERTI, S. 2015. Epigenetic inheritance and the missing heritability. *Hum Genomics,* 9**,** 17.

TRUSSLER, A. 2010. *The effect of a protein-restricted diet during pregnancy on the expression of the amino acid transporter System B0,+ in early rat embryos.* Master of Science, University of Waterloo.

TWINE, N. A., JANITZ, K., WILKINS, M. R. & JANITZ, M. 2011. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS One,* 6**,** e16266.

UAUY, R., KAIN, J. & CORVALAN, C. 2011. How can the Developmental Origins of Health and Disease (DOHaD) hypothesis contribute to improving health in developing countries? *Am J Clin Nutr,* 94**,** 1759S-1764S.

UEDA, P. 2013. *Famine in Nigeria and vitamin D in Sweden: two early life exposures and their relation to cardiovascular risk in adulthood.* PhD, Karolinska Institutet.

VALENCIA, S. H., OCAMPO, I. D., ARCE-PLATA, M. I., RECHT, J. & AREVALO-HERRERA, M. 2016. Glucose-6-phosphate dehydrogenase deficiency prevalence and genetic variants in malaria endemic areas of Colombia. *Malar J,* 15**,** 291.

VAMBERGUE, A. & FAJARDY, I. 2011. Consequences of gestational and pregestational diabetes on placental function and birth weight. *World J Diabetes,* 2**,** 196-203.

VASYLYEVA, T. L. & FERRY, R. J., JR. 2007. Novel roles of the IGF-IGFBP axis in etiopathophysiology of diabetic nephropathy. *Diabetes Res Clin Pract,* 76**,** 177-86.

VENET, D., DUMONT, J. E. & DETOURS, V. 2011. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol,* 7**,** e1002240.

VERHAEGHE, J., VAN BREE, R., VAN HERCK, E., LAUREYS, J., BOUILLON, R. & VAN ASSCHE, F. A. 1993. C-peptide, insulin-like growth factors I and II, and insulin-like growth factor binding protein-1 in umbilical cord serum: correlations with birth weight. *Am J Obstet Gynecol,* 169**,** 89-97.

VERMA, M. 2012. Epigenome-Wide Association Studies (EWAS) in Cancer. *Curr Genomics,* 13**,** 308-13.

VICKERS, M. H. 2011. Developmental programming of the metabolic syndrome - critical windows for intervention. *World J Diabetes,* 2**,** 137-48.

WAGNER, J. R., BUSCHE, S., GE, B., KWAN, T., PASTINEN, T. & BLANCHETTE, M. 2014. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol,* 15**,** R37.

WANG, S., MORTAZAVI, L. & WHITE, K. A. 2008. Higher-order RNA structural requirements and small-molecule induction of tombusvirus subgenomic mRNA transcription. *J Virol,* 82**,** 3864-71.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet,* 10**,** 57-63.

WATANABE, R. M. 2011. Inherited destiny? Genetics and gestational diabetes mellitus. *Genome Med,* 3**,** 18.

WATERLAND, R. A., KELLERMAYER, R., LARITSKY, E., RAYCO-SOLON, P., HARRIS, R. A., TRAVISANO, M., ZHANG, W., TORSKAYA, M. S., ZHANG, J., SHEN, L., MANARY, M. J. & PRENTICE, A. M. 2010. Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet,* 6**,** e1001252.

WHEATCROFT, S. B. & KEARNEY, M. T. 2009. IGF-dependent and IGF-independent actions of IGF-binding protein-1 and -2: implications for metabolic homeostasis. *Trends Endocrinol Metab,* 20**,** 153-62.

WHITTAKER, P. G., STEWART, M. O., TAYLOR, A., HOWELL, R. J. & LIND, T. 1990. Insulin-like growth factor 1 and its binding protein 1 during normal and diabetic pregnancies. *Obstet Gynecol,* 76**,** 223-9.

WHO 2013. Diagnostic Criteria and Classification of Hyperglycaemia First Detected in Pregnancy. Geneva.

WILLIAMS, L., SEKI, Y., VUGUIN, P. M. & CHARRON, M. J. 2014. Animal models of in utero exposure to a high fat diet: a review. *Biochim Biophys Acta,* 1842**,** 507-19.

XU, D. D., WEN, F. Q., LV, R. Y., ZHANG, M., CHEN, Y. S. & CHEN, X. W. 2016a. [Gene promoter methylation in glucose-6-phosphate dehydrogenase de fi ciency]. *Zhongguo Dang Dai Er Ke Za Zhi,* 18**,** 405-9.

XU, S. N., WANG, T. S., LI, X. & WANG, Y. P. 2016b. SIRT2 activates G6PD to enhance NADPH production and promote leukaemia cell proliferation. *Sci Rep,* 6**,** 32734.

XU, Y., OSBORNE, B. W. & STANTON, R. C. 2005. Diabetes causes inhibition of glucose-6-phosphate dehydrogenase via activation of PKA, which contributes to oxidative stress in rat kidney cortex. *Am J Physiol Renal Physiol,* 289**,** F1040-7.

YANG, D., FARIS, R., HIXSON, D., AFFIGNE, S. & ROGLER, C. E. 1996. Insulin-like growth factor II blocks apoptosis of N-myc2-expressing woodchuck liver epithelial cells. *J Virol,* 70**,** 6260-8.

YANG, X., HSU-HAGE, B., ZHANG, H., ZHANG, C., ZHANG, Y. & ZHANG, C. 2002. Women with impaired glucose tolerance during pregnancy have significantly poor pregnancy outcomes. *Diabetes Care,* 25**,** 1619-24.

YAU, S. W., AZAR, W. J., SABIN, M. A., WERTHER, G. A. & RUSSO, V. C. 2015. IGFBP-2 - taking the lead in growth, metabolism and cancer. *J Cell Commun Signal,* 9**,** 125-42.

YIN, Y. N., LI, X. L., TAO, T. J., LUO, B. R. & LIAO, S. J. 2014. Physical activity during pregnancy and the risk of gestational diabetes mellitus: a systematic review and meta-analysis of randomised controlled trials. *Br J Sports Med,* 48**,** 290-5.

ZAGHLOOL, S. B., AL-SHAFAI, M., AL MUFTAH, W. A., KUMAR, P., FALCHI, M. & SUHRE, K. 2015. Association of DNA methylation with age, gender, and smoking in an Arab population. *Clin Epigenetics,* 7**,** 6.

ZHANG, C. & NING, Y. 2011. Effect of dietary and lifestyle factors on the risk of gestational diabetes: review of epidemiologic evidence. *Am J Clin Nutr,* 94**,** 1975S-1979S.

ZHANG, C., RAWAL, S. & CHONG, Y. S. 2016. Risk factors for gestational diabetes: is prevention possible? *Diabetologia,* 59**,** 1385-90.

ZHANG, F. F., CARDARELLI, R., CARROLL, J., FULDA, K. G., KAUR, M., GONZALEZ, K., VISHWANATHA, J. K., SANTELLA, R. M. & MORABIA, A. 2011. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics,* 6**,** 623-9.

ZHANG, Z., APSE, K., PANG, J. & STANTON, R. C. 2000. High glucose inhibits glucose-6-phosphate dehydrogenase via cAMP in aortic endothelial cells. *J Biol Chem,* 275**,** 40042-7.

ZHANG, Z., LIEW, C. W., HANDY, D. E., ZHANG, Y., LEOPOLD, J. A., HU, J., GUO, L., KULKARNI, R. N., LOSCALZO, J. & STANTON, R. C. 2010. High glucose inhibits glucose-6-phosphate dehydrogenase, leading to increased oxidative stress and beta-cell apoptosis. *FASEB J,* 24**,** 1497-505.

ZHAO, S., FUNG-LEUNG, W. P., BITTNER, A., NGO, K. & LIU, X. 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One,* 9**,** e78644.

ZHU, C., YANG, H., GENG, Q., MA, Q., LONG, Y., ZHOU, C. & CHEN, M. 2015. Association of oxidative stress biomarkers with gestational diabetes mellitus in pregnant women: a case-control study. *PLoS One,* 10**,** e0126490.

ZHU, Y., MENDOLA, P., ALBERT, P. S., BAO, W., HINKLE, S. N., TSAI, M. Y. & ZHANG, C. 2016. Insulin-Like Growth Factor Axis and Gestational Diabetes Mellitus: A Longitudinal Study in a Multiracial Cohort. *Diabetes,* 65**,** 3495-3504.

ZOU, J., LIPPERT, C., HECKERMAN, D., ARYEE, M. & LISTGARTEN, J. 2014. Epigenome-wide association studies without the need for cell-type composition. *Nat Methods,* 11**,** 309-11.

ZOZULINSKA, D., MAJCHRZAK, A., SOBIESKA, M., WIKTOROWICZ, K. & WIERUSZ-WYSOCKA, B. 1999. Serum interleukin-8 level is increased in diabetic patients. *Diabetologia,* 42**,** 117-8.

APPENDIXES

APPENDIX Ai: Patient and control informed consent sheet

GENOME-WIDE GENE EXPRESSION ANALYSIS IN BLACK SOUTH AFRICAN WOMEN WHO DEVELOP GESTATIONAL DIABETES MELLITUS

Consent to Participate in a Genetic Research Study

By signing this form, you agree to participate in a research study that aims to identify genes involved in the development of gestational diabetes, This form is for you to give us permission to take two blood samples from you at your week 5 visit and a placental sample biopsy at delivery

By signing this form, I agree that:

- You have explained this study to me. You have answered all my questions

  YES ☐ NO ☐

- You have explained the possible harms and benefits (if any) of this study.

  YES ☐ NO ☐

- I understand that I have the right to refuse to take part in the study and the right to withdraw my participation at any time at no disadvantage to me.

  YES ☐ NO ☐

- I understand that no information about who I am will be given to anyone.

  YES ☐ NO ☐

- I will provide a blood sample for the genetic study and give permission for the use of a cord blood sample and placental biopsy as well.

  YES ☐ NO ☐

- These samples will be used as a source of DNA and RNA and stored indefinitely for research purposes.

  YES ☐ NO ☐

- I agree that a small bit of my DNA and RNA may be sent out of the country for research purposes.

  YES ☐ NO ☐

- I understand that every time a new study is done on my DNA and RNA, permission will be obtained from the ethics committee for the study to make sure that it is used only for the purposes stated above.

  YES ☐ NO ☐

I have read and understood the information sheet provided as well as this consent form. I agree, or consent, to take part in this study.

Full name of participant _____

_____ _____

Signature of participant Date

_____ _____

Signature of witness (or person who explained study and consent) Date

APPENDIX Aii: Patient and control information sheets

CASE INDIVIDUAL

GENOME-WIDE GENE EXPRESSION ANALYSIS IN BLACK SOUTH AFRICAN WOMEN WHO DEVELOP GESTATIONAL DIABETES MELLITUS

Information Sheet

Dear Potential Participant,

My name is Angela Hobbs and I am a PhD student at the University of the Witwatersrand. The results from your Oral Glucose Tolerance Test (OGTT) showed that your blood glucose levels are higher than what we would expect during pregnancy. I would therefore like to invite you to participate in a genetics study aimed at finding out what causes an increase in glucose levels during pregnancy.

Researchers Statement

We are inviting you to participate in a research study called "Genome-wide analysis of black South African women who develop gestational diabetes mellitus (GDM)". This study would like to examine which of our genes are involved in the development of this disease. The purpose of this information sheet is to provide you with the information you will need to make an informed decision about participating in this study. You may ask questions about the purpose of the study, what we require from you as a participant, the possible risks and benefits and anything else about the research project that may not be clear.

Purpose of Research

Genes are the parts of DNA which tell our bodies how to develop, grow and function. A change in the sequence of our DNA is called a mutation. Epigenetics is the study of changes in gene activity which are *not* caused by changes in the DNA sequence. The purpose of this study is to discover epigenetic changes that might cause gestational diabetes. When pregnant women show high blood glucose levels first seen during pregnancy, it is called gestational diabetes mellitus (GDM). This disease usually disappears after your baby has been born; however, it makes the mother and the fetus, more likely to develop Type 2 diabetes (T2D) later in life. Everyone has a number of epigenetic variations that result in the normal differences we expect to see between people, but some epigenetic changes lead to the development of diseases and problems with growth and development. We are trying to find out if there are epigenetic variations in specific genes in women who develop GDM in comparison to women who do not. This may help us better understand the role of genes in the development of GDM.

Description of the Research

You are already participating in a study investigating maternal factors associated with foetal growth and delivery outcomes.

- If you agree to participate in this genetics study, we would need to collect two blood samples from you at your week 5 visit.
- If you agree to participate, we will also analyse the genetic material extracted from the cord blood sample and placental biopsy sample which was taken directly after delivery.

Participation

Participation in this study is entirely voluntary and you will not be disadvantaged in any way if you choose not to participate. You and your family will not benefit from participating in this study and will not receive any remuneration. Although you may not benefit directly from this study, results from the study will improve the understanding of gestational diabetes mellitus and may benefit pregnant women in the future. If at any time you decide to withdraw your genetic material from participation in the study, you can do so (you DNA and RNA sample will then be destroyed).

Confidentiality

We will respect your privacy. No information about you or who you are will be given to anyone or published without your permission. The data from this study will be stored in a secured location and only members of the research team will have access to the data. Published study results will not reveal your identity. In the laboratory, confidentiality will be maintained at all times by assigning lab codes to the DNA and RNA sample. The DNA and RNA isolated from your blood will be stored indefinitely with these lab codes. Your samples will be stored in a secure laboratory at the Sydney Brenner Institute for Molecular Bioscience.

The test will involve no cost to the participant. Participation is completely voluntary and you will not be disadvantaged in any way if you decide not to participate. You may withdraw from the study at any time.

Should you require any further information, please do not hesitate to contact:

Mrs Angela Hobbs-Steyn or Professor Michele Ramsay
Division of Human Genetics, National Health Laboratory Service (NHLS) and the University of the Witwatersrand
(011) 489 9344 and (011) 489 9214

CONTROL INDIVIDUAL

GENOME-WIDE GENE EXPRESSION ANALYSIS IN BLACK SOUTH AFRICAN WOMEN WHO DEVELOP GESTATIONAL DIABETES MELLITUS

Information Sheet

Dear Potential Participant,

My name is Angela Hobbs and I am a PhD student at the University of the Witwatersrand. The results from your Oral Glucose Tolerance Test (OGTT) showed that your blood glucose levels are normal. We need control samples for a research study and would therefore like to invite you to participate.

Researchers Statement

We are inviting you to participate in a research study called "Genome-wide analysis of black South African women who develop gestational diabetes mellitus (GDM)". This study would like to examine which of our genes are involved in the development of this disease. The purpose of this information sheet is to provide you with the information you will need to make an informed decision about participating in this study. You may ask questions about the purpose of the study, what we require from you as a participant, the possible risks and benefits and anything else about the research project that may not be clear.

Purpose of Research

Genes are the parts of DNA which tell our bodies how to develop, grow and function. A change in the sequence of our DNA is called a mutation. Epigenetics is the study of changes in gene activity which are *not* caused by changes in the DNA sequence. The purpose of this study is to discover epigenetic changes that might cause gestational diabetes. When pregnant women show high blood glucose levels first seen during pregnancy, it is called gestational diabetes mellitus (GDM). This disease usually disappears after your baby has been born; however, it makes the mother and the fetus, more likely to develop Type 2 diabetes (T2D) later in life. Everyone has a number of epigenetic variations that result in the normal differences we expect to see between people, but some epigenetic changes lead to the development of diseases and problems with growth and development. We are trying to find out if there are epigenetic variations in specific genes in women who develop GDM in comparison to women who do not. This may help us better understand the role of genes in the development of GDM.

Description of the Research

You are already participating in a study investigating maternal factors associated with foetal growth and delivery outcomes.

- If you agree to participate in this genetics study, we would need to collect two blood samples from you at your week 5 visit.
- If you agree to participate, we will also analyse the genetic material extracted from the cord blood sample and placental biopsy sample which was taken directly after delivery.

Participation

Participation in this study is entirely voluntary and you will not be disadvantaged in any way if you choose not to participate. You and your family will not benefit from participating in this study and will not receive any

remuneration. Although you may not benefit directly from this study, results from the study will improve the understanding of gestational diabetes mellitus and may benefit pregnant women in the future. If at any time you decide to withdraw your genetic material from participation in the study, you can do so (you DNA and RNA sample will then be destroyed).

Confidentiality

We will respect your privacy. No information about you or who you are will be given to anyone or published without your permission. The data from this study will be stored in a secured location and only members of the research team will have access to the data. Published study results will not reveal your identity. In the laboratory, confidentiality will be maintained at all times by assigning lab codes to the DNA and RNA sample. The DNA and RNA isolated from your blood will be stored indefinitely with these lab codes. Your samples will be stored in a secure laboratory at the Sydney Brenner Institute for Molecular Bioscience.

The test will involve no cost to the participant. Participation is completely voluntary and you will not be disadvantaged in any way if you decide not to participate. You may withdraw from the study at any time.

Should you require any further information, please do not hesitate to contact:

Mrs Angela Hobbs-Steyn or Professor Michele Ramsay
Division of Human Genetics, National Health Laboratory Service (NHLS) and the University of the Witwatersrand
(011) 489 9344 and (011) 489 9214

APPENDIX Bi: Ethics clearance certificate

M130420

R14/49 Ms Angela Hobbs

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M130420

NAME: (Principal Investigator)	Ms Angela Hobbs
DEPARTMENT:	School of Pathology/Human Genetics National Health Laboratory Services
PROJECT TITLE:	Genome-Wide Gene Expression Analysis in Black South African women Who Develop Gestational Diabetes Mellitus
DATE CONSIDERED:	26/04/2013
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Prof Michele Ramsay
APPROVED BY:	Professor PE Cleaton-Jones, Chairperson, HREC (Medical)

DATE OF APPROVAL: 22/07/2013

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Secretary in Room 10004, 10th floor, Senate House, University.
I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**.

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX Bii: Amendment to ethics certificate

Human Research Ethics Committee (Medical)

Research Office Secretariat: Senate House Room SH 10005, 10th floor. Tel +27 (0)11-717-1252
Medical School Secretariat: Phillip Tobias Building. 2nd Floor Tel +27 (0)11-717-2700
Private Bag 3, Wits 2050, www.wits.ac.za. Fax +27 (0)11-717-1265

18 November 2014

Ms Angela Hobbs
Division of Human Genetics
National Health Laboratory Services
P.O Box 1038
Johannesburg
2000
Sent by email to: ang.hobbs15@gmail.com

Dear Ms Hobbs

Re: Protocol Ref no: M130420
Protocol Title: Genome wide Gene expression Analysis in Black South African Women who develop Gestational Diabetes Mellitus
Principal Investigator: Angela Hobbs

Protocol Amendment

This letter serves to confirm that the Chairman of the Human Research Ethics Committee (Medical) has approved the amendments on the above mentioned protocol, as detailed in your letter received on 03 October 2014.

The following documents were received:

- Research proposal

Thank you for keeping us informed and updated,

Yours Sincerely,

.....

Mr Langutani Masingi
Administrative Officer
Human Research Ethics Committee (Medical)

UNIVERSITY OF THE WITWATERSRAND
OFFICE OF THE DEPUTY VICE CHANCELLOR RESEARCH
18 NOV 2014
RESEARCH OFFICE
JOHANNESBURG

APPENDIX Ci: Tempus™ Blood RNA Tube and Tempus™ Spin RNA Isolation Kit (Applied Biosystems)

1. If the sample is frozen, thaw the sample in the Tempus tube at room temperature (18 to 25 °C).

2. Remove the cap from the Tempus tube, then pour the contents of the tube into a clean 50-mL.

3. Pipet 3 mL of 1✕ PBS (Ca2+/Mg2+-free) into the tube to bring the total volume to 12 mL. Note: If the initial blood sample was less than 3 mL, make up the difference by adding enough 1✕ PBS to bring the total volume to 12 mL. Otherwise, RNA yields decrease significantly.

4. Replace the cap on the tube, then vortex the tube *vigorously* (at maximum vortex speed) for 30 seconds to ensure proper mixing of the contents. Note: To prevent the tube from leaking and spraying the sample during vortexing, make sure the tube is capped properly. Vortex the diluted sample for at least 30 seconds; vortexing for less than 30 seconds may cause clogging of the purification consumable.

5. Centrifuge the tube at 4 °C at 3,000 x *g* (rcf) for 30 minutes.

6. Carefully pour off the supernatant. Note: The RNA pellet is transparent and invisible, Handle the tube carefully so that you do not shake the RNA pellet off the bottom of the tube.

7. Leave the tube inverted on absorbent paper for 1 to 2 minutes.

8. Blot the remaining drops of liquid off the rim of the tube with clean absorbent paper.

9. Pipet 400 µL of RNA Purification Resuspension Solution into the tube, then vortex briefly to resuspend the RNA pellet. Note: To prevent washing any blood residue down the inside of the tube, insert the pipet tip into the tube and add the resuspension solution to the bottom of the tube.

10. The resuspended RNA can be kept on ice while preparing for the next steps: Proceed to "Performing the Purification Run"

To perform the purification run:
1. Label the RNA purification filter, then insert the filter into a waste collection tube.

2. Pre-wet the filtration membrane by pipeting RNA Purification Wash Solution 1 into the purification filter.

3. Pipet the resuspended RNA into the purification filter, then centrifuge.

4. Remove the purification filter, discard the liquid waste collected in the waste tube, then re-insert the purification filter into the waste tube. **Note:** Each time you discard the liquid waste, instead of reusing the waste tube, you can transfer the purification filter into a new collection tube.

5. Pipet RNA Purification Wash Solution 1 into the purification filter, then centrifuge.

6. Remove the purification filter, discard the liquid waste collected in the waste tube, then re-insert the purification filter into the waste tube. Pipet RNA Purification Wash Solution 2 into the purification filter, then centrifuge. Note: When a DNase treatment is required, extend the centrifuge time to 1 minute to remove all wash solutions and dry the membrane completely.

8. Remove the purification filter, discard the liquid waste collected in the waste tube, then re-insert the purification filter into the waste tube.

10. Pipet RNA Purification Wash Solution 2 into the purification filter, then centrifuge.

11. Remove the purification filter, discard the liquid waste collected in the waste tube, then re-insert the purification filter into the waste tube. Centrifuge to dry the membrane.

12. Transfer the purification filter to a new, labelled collection tube to collect the eluate.

13. Pipet Nucleic Acid Purification Elution Solution into the purification filter, close the cap, incubate the entire tube, then centrifuge.

14. Pipet the collected RNA eluate back into the purification filter, then centrifuge. No incubation is necessary.

15. Discard the purification filter, then transfer approximately 90 µL of the RNA eluate to a new, labelled collection tube. **Note**: When transferring the RNA eluate, carefully pipet the liquid out of the collection tube starting from the top of the liquid to ensure that the pelleted particulates are not disturbed.

16. Replace the cap on the new collection tube, then store the RNA at −20 ℃, or −80 ℃ for long-term storage.

APPENDIX Cii: Salting out method (Miller et al., 1988)

Equipment and Materials

- Polypropylene tubes 15ml
- Lysis buffer (10mM Tris-HCL,400mM NaCl, 2mM Na_2EDTA, pH 8.2)
- SDS 10%
- Proteinase K solution (1 mg proteinase K in 1% SDS and 2 mM Na_2 EDTA),
- Centrifuge
- Absolute ethanol
- TE buffer (10mM Tris-HCL, 0.2mM Na_2 EDTA, pH 7.5)
- Disposable gloves
- Gilson pipette

Procedure

1. Resuspend the buffy coats of nucleated cells obtained from blood with anticoagulents (ACD or EDTA) with 3ml of nuclear lysis buffer.
2. Digest the cell lysates, with 0.2 ml of 10% SDS and 0.5 ml of proteinase K solution, overnight at 37 °C.
3. Add 1ml of saturated NaCl (6M) to each tube and shake vigorously for 15 seconds.
4. Centrifuge for 15 minutes at 2500 rpm, 4°C
5. Transfer the supernatant containing the DNA to another 15ml polypropylene tube, the precipitated protein pellet is left behind at the bottom of the tube.
6. Add 2 volumes of absolute ethanol and invert the tubes several times until the DNA precipitates.
7. Remove the precipitated DNA with a plastic spatula or pipette and transfer to a 1.5ml microcentrifuge tube containing 100-200 µl TE buffer.
8. Dissolve the DNA for 2 hours at 37°C.
9. Store the tube at +4 or –20°C.

Check quantity/quality of DNA (see QUALITY CONTROL OF DNA protocol

APPENDIX Ciii: RNeasy Mini Kit (Qiagen)– Purification of Total RNA from Animal Tissues

Procedure

1. Remove the RNA*later* stabilized tissue sample from the reagent using forceps. Weigh the tissue. Do not use more than 30 mg.

2. Place the 30 mg tissue directly into a suitably sized vessel for disruption and homogenization.

3. Disrupt the tissue and homogenize the lysate in 600 µl Buffer RLT (do not use more than 30 mg tissue) using a TissueLyser.

4. Centrifuge the lysate for 3 min at full speed. Carefully remove the supernatant by pipetting, and transfer it to a new microcentrifuge tube (not supplied). Use only this supernatant (lysate) in subsequent steps.

5. Add 1 volume of 70% ethanol to the cleared lysate, and mix immediately by pipetting. Do not centrifuge. Proceed immediately to step 6.
Note: The volume of lysate may be less than 600 µl due to loss during homogenization and centrifugation in steps 3 and 4.

6. Transfer up to 700 µl of the sample, including any precipitate that may have formed, to an RNeasy spin column placed in a 2 ml collection tube (supplied). Close the lid gently, and centrifuge for 15 s at ≥8000 x *g* (≥10.000 rpm). Discard the flow-through.

7. Add 700 µl Buffer RW1 to the RNeasy spin column. Close the lid gently, and centrifuge for 15 s at ≥8000 x *g* (≥10.000 rpm) to wash the spin column membrane. Discard the flow-through.
Note: After centrifugation, carefully remove the RNeasy spin column from the collection tube so that the column does not contact the flow-through. Be sure to empty the collection tube completely.

8. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid gently and centrifuge for 15 s at ≥8000 x *g* (≥10.000 rpm) to wash the spin column membrane. Discard the flow-through.
Note: Buffer RPE is supplied as a concentrate. Ensure that ethanol is added to Buffer RPE before use.

9. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid gently. and centrifuge for 2min at ≥8000 x *g* (≥10.000 rpm) to wash the spin column membrane.
The long centrifugation dries the spin column membrane ensuring that no ethanol is carried over during RNA elution. Residual ethanol may interfere with downstream reactions. **Note**: After centrifugation, carefully remove the RNeasy spin column from the collection tube so that the column does not contact the flow-through. Otherwise. carryover of ethanol will occur.

10. Optional: Place the RNeasy spin column in a new 2 ml collection tube (supplied). and discard the old collection tube with the flow-through. Close the lid gently, and centrifuge at full speed for 1 min.
Perform this step to eliminate any possible carryover of Buffer RPE, or if residual flow-through remains on the outside of the RNeasy spin column after step 9.

11. Place the RNeasy spin column in a new 1.5 ml collection tube (supplied). Add 30–50 µl RNase-free water directly to the spin column membrane. Close the lid gently, and centrifuge for 1 min at ≥8000 x *g* (≥10.000 rpm) to elute the RNA.

12. If the expected RNA yield is >30 µg, repeat step 11 using another 30–50 µl RNasefree water, or using the eluate from step 11 (if high RNA concentration is required). Reuse the collection tube from step 11.
If using the eluate from step 11, the RNA yield will be 15–30% less than that obtained using a second volume of RNase-free water, but the final RNA concentration will be higher.

APPENDIX Civ: DNeasy Blood and Tissue kit (Qiagen)

1. **Tissue: Cut tissue** ≤25 mg of the tissue into small pieces. and place in a 1.5 ml microcentrifuge tube.

2. **Add 200 µl *Buffer AL***. Mix thoroughly by vortexing and incubate the samples at 56°C for 10 min.

3. **Add 200 µl *ethanol*** (96–100%). Mix thoroughly by vortexing and pipet the mixture into a DNeasy Mini spin column placed in a 2 ml collection tube. Centrifuge at ≥ 6000 x *g* (8000 rpm) for 1 min. Discard the flow-through and collection tube. Place the spin column in a new 2 ml collection tube.

4. **Add 500 µl *Buffer AW1***. Centrifuge for 1 min at ≥6000 x *g*. Discard the flow-through and collection tube. Place the spin column in a new 2 ml collection tube.

5. **Add 500 µl *Buffer AW2***, and centrifuge for 3 min at 20.000 x *g* (14.000 rpm). Discard the flow-through and collection tube. Transfer the spin column to a new 1.5 ml or 2 ml microcentrifuge tube.

6. **Elute the DNA** by adding 200 µl ***Buffer AE*** to the centre of the spin column membrane. Incubate for 1 min at room temperature (15–25°C). Centrifuge for 1 min at ≥6000 x *g*.

APPENDIX D: GLOBINclear™ Kit (Ambion)

1. Reagent Preparation
Add 2 mL of 100% isopropanol to the RNA Binding Buffer Concentrate. Mix well and mark the label to indicate that the isopropanol was added. Store at room temperature. This mixture is referred to as ***RNA Binding Buffer.***

Add 4 mL of 100% ethanol to the RNA Wash Solution Concentrate. Mix well and indicate on the label that the ethanol was added. Store at room temperature. The resulting mixture is referred to as ***RNA Wash Solution*** in the instructions.

Dilute the RNA Binding Beads in RNA Bead Buffer, and add isopropanol. In a 1.5 mL tube, combine RNA Bead Buffer (10 µl/sample) with RNA Binding Beads and mix briefly. Add the 100% isopropanol and mix thoroughly by vortexing. Store at room temperature. This mixture is the ***Bead Resuspension Mix***

2. Preparation of Streptavidin Magnetic Beads
Set a dry incubator to 50°C and warm the 2X Hybridization Buffer and the Streptavidin Bead Buffer to 50°C for at least 15 min before starting the next procedure. Vortex well before use.

a. Place 30 µL of Streptavidin Magnetic Beads per sample into a 1.5 mL tube
i. Use 30 µL of Streptavidin Magnetic Beads for each sample; calculate the volume of beads needed for the samples being processed that day. When there are more than 2 samples, it is prudent to include 5–10% overage to cover pipetting error.
ii. Vortex the tube of the Streptavidin Magnetic Beads to suspend the settled beads, and transfer the volume needed into a 1.5 mL Non-stick Tube
iii. Briefly centrifuge (<2 sec) at low speed (<1000 x g) to collect the mixture at the bottom of the tube.

b. Magnetically capture the beads and carefully remove and discard the supernatant
i. Place the tube on a magnetic stand to capture the Streptavidin Magnetic Beads. Leave the tube on the magnetic stand until the mixture becomes transparent (~3–5 min). indicating that capture is complete.
ii. Carefully aspirate the supernatant using a pipet without disturbing the Streptavidin Magnetic Beads. Discard the supernatant, and remove the tube from the magnetic stand

c. Equilibrate the beads with an equal volume of Streptavidin Bead Buffer and place at 50°C
i. Add Streptavidin Bead Buffer to the Streptavidin Magnetic Beads; use a volume equal to the original volume of Streptavidin Magnetic Beads. Vortex vigorously until the beads are resuspended.
ii. Place the prepared Streptavidin Magnetic Beads at 50°C and immediately proceed to the next step. The beads should remain at 50°C for at least 15 min before they are used

3. Hybridization of Globin mRNA and Globin Capture Oligonucleotides
a. Combine 1–10 µg RNA and 1 µL Capture Oligo Mix
Combine the following in a 1.5 mL Non-stick Tube provided with the kit: 1–10 µg human blood total RNA (in a maximum volume of 14 µL. i.e. the RNA concentration must ≥70 ng/µL) and add 1uL of Capture Oligo Mix

b. Add Nuclease-free Water for a final volume of 15 µL
If necessary, add Nuclease-free Water to the sample mixture from step 1 to a final volume of 15 µL.

c. Add 15 µL 2X Hybridization Buffer
a. Add 15 µL of 50°C 2X Hybridization Buffer to the sample.
b. Vortex briefly to mix and centrifuge briefly at low speed to collect the contents in the bottom of the tube.

4. Hybridize at 50°C for 15 min
Place the sample in a prewarmed 50°C incubator and allow the Globin Capture Oligo Mix to hybridize to the globin mRNA for 15 min.

5. Removal of Globin mRNA

a. Add 30 µL prepared Streptavidin Magnetic Beads to each sample

i. Remove the prepared Steptavidin Magnetic Beads from the 50°C incubator, and resuspend them by gentle vortexing. Briefly centrifuge (<2 sec) at low speed (<1000 x g) to collect the mixture at the bottom of the tube.
ii. Add 30 µL of prepared Streptavidin Magnetic Beads to each RNA sample.
iii. Vortex to mix well and centrifuge briefly at low speed as in the previous steps to collect the contents in the bottom of the tube.
iv. Flick the tube very gently to resuspend the beads, being careful to keep the contents at the bottom of the tube.

b. Incubate 30 min at 50°C Place the RNA bead mixture at 50°C (hybridization oven or other fixed temperature air incubator recommended) and incubate for 30 min.

c. Magnetically capture the beads

i. Remove sample from the incubator, and vortex briefly to mix. Centrifuge briefly at low speed to collect the contents in the bottom of the tube.
ii. Capture the Streptavidin Magnetic Beads on a magnetic stand. Leave the tube on the magnetic stand until the mixture becomes transparent (~3–5 min). indicating that capture is complete.

d. Transfer the supernatant containing the RNA to a new tube

Carefully draw up the supernatant. which contains the globin mRNA depleted RNA, using a pipet without disturbing the Streptavidin Magnetic Beads. Transfer the RNA to a new 1.5 mL Non-stick Tube supplied with the kit, and place on ice. ***The supernatant contains the*** GLOBINclear ***RNA; do not discard the supernatant.***

6. Purify the GLOBINclear RNA

Warm the Elution Buffer to 58°C.

a. Add 100 µL RNA Binding Buffer to each sample

b. Add 20 µL *Bead Resuspension Mix* to each sample; mix for 10 sec

i. Vortex the ***Bead Resuspension Mix***. then immediately dispense 20 µL to each sample. It is important to resuspend the beads thoroughly before adding them to the samples.
ii. Vigorously vortex the sample for 10 sec to fully mix the reagents, and to allow the RNA Binding Beads to bind the RNA.
iii. Briefly centrifuge (<2 sec) at low speed (<1000 x g) to collect the mixture at the bottom of the tube.

c. Magnetically capture the RNA Binding Beads and discard the supernatant

i. Capture the RNA Binding Beads by placing the tube on a magnetic stand. Leave the tube on the magnetic stand until the mixture becomes transparent (~3–5 min), indicating that capture is complete.
ii. Carefully aspirate the supernatant using a pipet without disturbing the RNA Binding Beads. Discard the supernatant.
iii. Remove the tube from the magnetic stand. It is critical for effective washing to remove the tube from the magnetic stand before adding the RNA Wash Solution

d. Wash the RNA Binding Beads with 200 µL RNA Wash Solution

i. Add 200 µL RNA Wash Solution to each sample and vortex for 10 sec.
ii. Briefly centrifuge (<2 sec) at low speed (<1000 x g) to collect the mixture at the bottom of the tube.
iii. Capture the RNA Binding Beads on a magnetic stand as in the previous magnetic bead capture steps.
iv. Carefully aspirate and discard the supernatant, and remove the tube from the magnetic stand.

7. Purify the GLOBINclear RNA

a. Remove any remaining supernatant and leave the tube open for 5 min
i. Briefly centrifuge the tube as in previous steps and place it back on the magnetic stand.
ii. Remove any liquid in the tube with a small-bore pipet tip
iii. Remove the tube from the magnetic stand and allow the beads to air-dry for 5 min with the caps left open.

b. Add 30 µL Elution Buffer and incubate at 58°C for 5 min to elute the enriched RNA
i. Add 30 µL warm (58°C) Elution Buffer to each sample, and vortex vigorously for ~10 sec to thoroughly resuspend the RNA Binding Beads.
ii. Incubate the mixture at 58°C for 5 min.
iii. Vortex the sample vigorously for ~10 sec to thoroughly resuspend the RNA Binding Beads and centrifuge briefly at low speed as in previous steps to collect the mixture at the bottom of the tube.

8. Magnetically capture the RNA Binding Beads and transfer the GLOBINclear RNA to a new tube
a. Capture the RNA Binding Beads on a magnetic stand as in the previous magnetic bead capture steps. Be especially careful at this step to avoid disturbing the RNA Binding Beads when collecting the supernatant. ***The purified RNA will be in the supernatant.***
b. Transfer the supernatant containing the RNA to a 1.5 mL Non-stick Tube (supplied). Store the purified RNA at –20°C.

APPENDIX E: TruSeq Stranded mRNA Preparation Protocol (Illumina)

Purify and Fragment mRNA

- Illumina recommends the use of 0.1-4ug RNA
- Remove the following from the freezer and thaw at room temperature
- Bead Binding Buffer (BBB)
- Bead washing Buffer (BWB)
- Elution buffer (ELB)
- Fragment, Prime and Finish mix (FPF)
- Resuspension Buffer (RSB)
- Remove the RNA purification beads from the fridge and thaw to room temperature
- Pre-program the PCR machine to the following programmes:
    - 65ºC 5 minutes, 4ºC hold (**mRNA degradation**)
    - 80ºC 2 minutes, 25ºC hold (**mRNA elution 1**)
    - 94ºC 8 minutes, 4ºC hold (**Elution 2-Frag-Prime**)
- Centrifuge should be at 25ºC
- Apply RBP barcode to plate

Make RBP

1. Dilute RNA in nuclease-free ultra pure water to 50ul in a 96-well plate. Vortex the RNA purification beads vigorously to resuspend
2. Add 50ul RNA purification beads to each well, gently pipette up and down 6 times. Seal plate

Incubate 1 RBP

1. Place sealed plate into PCR machine, choose **mRNA denaturation.** Remove plate when it reached 4ºC
2. Incubate plate at room temperature for 5 minutes

Wash RBP

1. Remove seal from plate. Place on magnetic stand for 5 minutes. Remove and discard all supernatant form each well. Remove plate from magnetic stand
2. Wash beads with 200ul BWB, gently pipette up and down 6 times. Place plate on magnetic stand for 5 minutes. Centrifuge thawed ELB to 600g for 5 seconds
3. Remove and discard all supernatant from each well. Remove plate from magnetic stand
4. Add 50ul ELB to each well, gently pipette up and down 6 times. Seal plate. Store ELB at 4ºC

Incubate 2 RBP

1. Place sealed plate in PCR machine, choose **mRNA Elution 1.** Remove plate when reached 25ºC
2. Incubate at room temperature for 5 minutes. Remove seal

Make RFP

1. Centrifuge thawed BBB to 600g for 5 seconds. Add 50ul BBB to each well, gently pipette up and down 6 times. Incubate at room temperature for 5 minutes, store BBB in the fridge
2. Place plate on magnetic stand for 5 minutes. Remove and discard all the supernatant. Remove plate from magnetic stand

3. Wash beads with 200ul BWB, gently pipette up and down 6 times. Store BWB in the fridge. Place plate on magnetic stand for 5 minutes
4. Remove and discard supernatant. Remove plate from magnetic stand. Add 19,5ul FPF, gently pipette up and down 6 times
5. Seal plate, store FPF in feezer

Incubate RFP

1. Place plate in PR machine and choose **Elution 2-Frag-Prime.** Remove plate when reached 4ºC
2. Proceed *immediately* to next step

Synthesize First strand cDNA

- Remove First Strand Synthesis Act D Mix (FSA) from freezer and thaw at room temperature
- Programme PCR machine to the following:
- **Synthesize 1st Strand**
    - Pre-heat lid to 100ºC
    - 25ºC for 10 minutes
    - 42ºC for 15 minutes
    - 70ºC for 15 minutes
    - Hold for 4ºC
- Apply CDP barcode to new plate

Make CDP

1. Remove seal from RBP plate. Place plate on magnetic stand fro 5 minutes
2. Transfer 17ul of supernatant from each well of plate to a new (CDP plate) . Centrifuge the FSA to 600g for 5 seconds
3. Add 50ul SuperScript II to the FSA tube (otherwise add SuperScript II in a ratio of 1ul for each 9ul FSA) centrifuge briefly. Add 8ul of the mix to each well, gently pipette up and down 6 times
4. Seal plate, centrifuge briefly. Store mix in the freezer

Incubate 1 CDP

1. Place plate in PCR, choose **Synthesize 1st Strand.** When plate reached 4ºC remove and proceed *immediately* to next step

Synthesize Second Strand cDNA

- Thaw these reagents at room temperature: End Pair Control (CTE) and Second Strand Marking Master Mix (SMM). Resuspention Buffer (RSB) from the fridge. AMPure XP beads from fridge
- Pre-heat PCR machine to 16ºC
- Choose the pre-heat lid to 30ºC
- Apply ALP barcode to new PCR plate

Add SMM

1. Remove seal from CDP plate. Add 5ul RSB to each well. Centrifuge SMM to 600g for 5 seconds
2. Add 20ul SMM, gently pipette up and down 6 times. Seal plate, return SMM to freezer

Incubate 2 CDP

1. Place plate in PCR machine, incubate at 16ºC for 1 hour. Remove plate and place on bench, remove seal and bring plate to room temperature

Purify CDP

1. Vortex AMPure XP beads. Add 40ul AMPure XP beads, gently pipette up and down 10 times
2. Incubate at room temperature for 15 minutes. Place plate on magnetic stand for 5 minutes
3. Remove and discard 135ul supernatant
4. Add 200ul 80% EtOH without disturbing beads. Incubate for 30 seconds and remove and discard supernatant. Repeat for a total of two ethanol washes
5. Let the plate dry at room temperature for 15 minutes, remove plate from magnetic stand
6. Centrifuge RSB to 600g for 5 seconds. Add 17,5ul RSB, gently pipette up and down 10 times
7. Incubate plate at room temperature for 2 minutes. Place plate of magnetic stand for 5 minutes
8. Transfer 15ul supernatant (ds cDNA) from the plate to new plate labeled ALP

· SAFE STOPPING POINT : ***Store ALP plate at -15ºC to -25ºC for up to 7 days***

Adenylate 3' ends

- Remove the following and thaw at room temperature:
    - A-tailing control (CTA)
    - A-tailing mix (ATL)
    - Resuspension Buffer (RSB) from fridge
    - ALP plate, centrifuge once thawed
- Programme PCR machine as **ATAIL70**:
    - Pre-heat lid to 100ºC
    - 37ºC for 30 minutes
    - 70ºC for 5 minutes
    - Hold for 4ºC

Add ATL

1. Add 2.5ul RSB. Add 12.5ul ATL, gently pipette up and down 10 times. Seal the plate

Incubate 1 ALP

1. Place plate in PCR machine, choose **ATAIL70.** When reached 4ºC remove plate and proceed *immediately* to next step

Ligate Adapters

- Remove the following and thaw at room temperature:
    - Appropriate RNA adapter tubes (adapter indices being used)
    - Stop Ligation Buffer (STL)
    - Ligation control (CTL)
    - Resuspension Buffer (RSB)
    - AMPure XP beads
- Preheat PCR machine to 30ºC with preheat lid option to 100ºC
- Apply CAP barcode to new plate
- Apply PCR barcode to new plate

Add LIG

1. Centrifuge RNA adapter tubes at 600g for 5 seconds. Centrifuge STL at 600g for 5 seconds
2. Immediately before use remove Ligation Mix (LIG) from freezer. Remove seal from ALP plate
3. Add 2.5ul RSB to each well. Add 2.5ul LIG to each well, return to freezer immediately. Add 2.5ul of RNA adapter index to each well, gently pipette up and down 10 times
4. Seal the plate. Centrifuge ALP plate at 280g for 1 minute

Incubate 2 ALP

1. Place sealed ALP plate in PCR machine and incubate at 30ºC for 10 minutes. Remove plate from PCR machine

Add STL

1. Remove seal. Add 5ul STL to each well, seal plate
2. Place on shaker for 2 minutes at 1800rpm. Centrifuge ALP plate at 280g for 1 minute

Clean up ALP

1. Vortex AMPure XP beads, remove seal from plate. Add 42ul AMPure XP beads, place on shaker for 2 minutes at 1800rpm. Centrifuge ALP plate at 280g for 1 minute
2. Incubate at room temperature for 15 minutes. Place ALP plate on magnetic stand for 5 minutes
3. Remove and discard 79.5ul supernatant. Add 200ul 80% EtOH without disturbing the beads
4. Incubate at room temperature for 30 seconds and remove the supernatant. Repeat for a total of two ethanol washes
5. Let plate air dry for 15 minutes
6. Remove the ALP plate from the magnetic stand. Add 52.5ul RSB, place on shaker for 2 minutes at 1800rpm. Incubate at room temperature for 2 minutes
7. Centrifuge ALP plate at 280g for 1 minute. Place plate on magnetic stand for 5 minutes
8. Transfer 50ul of supernatant from ALP plate to CAP plate
9. Vortex AMPure XP beads. Add 50ul AMPure XP beads to CAP plate. Incubate CAP plate for 15 minutes. Centrifuge ALP plate at 280g for 1 minute
10. Place CAP plate on magnetic stand for 5 minutes. Remove and discard the 95ul supernatant form the CAP plate. Add 200ul 80% EtOH, do not disturb the beads
11. Incubate at room temperature for 30 seconds. Repeat for a total of two ethanol washes
12. Let plate air dry for 15 minutes, remove plate from magnetic stand
13. Add 22.5ul RSB. Incubate at room temperature for 2 minutes. Centrifuge ALP plate at 280g for 1 minute. Place plate on magnetic stand for 5 minutes
14. Transfer 20ul supernatant to PCR plate

· **SAFE STOPPING POINT :** ***PCR plate can be stored at -15ºC to -25ºC for up to 7 days***

Enrich DNA Fragments

- Remove and thaw at room temperature:
    - PCR Master Mix (PMM) and PCR Primer Cocktail (PPC)
    - Resuspension Buffer (RSB)
    - AMPure XP beads
    - Remove PCR plate if stored (centrifuge and remove seal)
- Programme PCR machine store as **PCR**:
    - Preheat lid to 100ºC
    - 98ºC for 30 seconds

    - 98ºC for 10 seconds
    - 60ºC for 30 seconds
    - 72ºC for 30 seconds

    } 15 cycles

    - 72ºC for 5 minutes
    - 4ºC for hold
- Apply TSP1 barcode to plate

Make PCR

1. Add 5ul PPC and 25ul PMM to each well, place on shaker for 2 minutes at 1800rpm. centrifuge PCR plate at 280g for 1 minute

Amp PCR

1. Place plate in PCR machine, choose **PCR**
2. Vortex AMPure XP beads. Add 50ul AMPure XP beads to each well of the plate containing the library, place on shaker for 2 minutes at 1800rpm
3. Incubate plate at room temperature for 15 minutes, centrifuge PCR plate at 280g for 1 minute
4. Place plate on magnetic stand for 5 minutes. Remove and discard the 95ul supernatant
5. Add 200ul 80% EtOH without disturbing beads. Incubate at room temperature for 30 seconds, remove and discard supernatant
6. Repeat for a total of two ethanol washes and let the plate air dry for 15 minutes
7. Add 32.5ul RSB, place on shaker for 2 minutes at 1800rpm. Incubate plate at room temperature for 2 minutes. centrifuge PCR plate at 280g for 1 minute. Place plate on magnetic stand
8. Transfer 30ul supernatant to TSP1 plate

· SAFE STOPPING POINT : ***TSP1 plate can be stored at -15ºC to -25ºC for up to 7 days***

Validate Library

- Illumina recommend quantifying the libraries using qPCR and to check quality using the Bioanalyser

Normalise and Pool Libraries

- Remove the TSP1 plate to thaw at room temperature and centrifuge
- Apply DCT barcode to plate. apply PDP barcode to plate

Make DCT

1. Transfer 10ul sample from TSP1 plate to DCT plate
2. Normalise concentration of each sample to 10nM using Tris-HCl 10mM, pH 8.5 with 0.1% Tween 20
3. Gently pipette up and down 10 times
4. For non-pooled libraries the protocol stops here. for pooled libraries proceed to next step

Make PDP (Pooling only)

1. Determine number of samples to be combined
2. If pooling 2-24 samples:
    a. Transfer 10ul of normalized samples into PDP plate
    b. The volume in each well is 10x the number of sample there in (20-240ul)
3. If pooling 25-96 samples:
    a. Transfer 5ul of normalized sample into PDP plate

    b. Repeat for the number of samples
4. Gently pipette up and down 10 times
5. Proceed to cluster generation
6. Seal the PDP plate

APPENDIX F: TaqMan Gene Expression Assays

A. Prepare the cDNA sample with the High Capacity RNA to cDNA kit (Applied Biosystems)

1. Use 250ng of total RNA
2. Allow the kit components to thaw on ice
3. Prepare a +RT and –RT reaction following the table below

Component	+RT reaction (ul)	-RT reaction (ul)
RNA sample	Up to 9	Up to 9
2x RT Buffer	10	10
2x Enzyme mix	1	
Nuclease-free ddH_20	Up to 20	Up to 20
Total/reaction	**20**	**20**

4. Aliquot reaction mix into a plate or tubes and seal properly
5. Centrifuge briefly and put on ice
6. Incubate at 37°C for 60 minutes; 95°C for 5 minutes; 4°C hold

B. Setting up the Real-time PCR assay

1. Thaw the reagents on ice (20X gene expression assay and cDNA samples (if frozen))
2. Mix the reagents by vortexing gently for a few seconds and centrifuge briefly
3. Calculate the number of reaction to prepare (run each sample in triplicate; NTC and Housekeeping genes)
4. For each sample, pipette the following into a 1.5ml Eppendorf tube

	Volume per 20ul reaction (ul)	
PCR reaction mix component	**Single reaction**	**Master mix (per assay)(80x)**
20X Taqman Gene Expression Assay	1	80
2x Taqman GeneEx master mix	10	800
cDNA (1 – 100ng)	1	80
RNAse-free water	8	640

5. Mix the master mix well by inverting and centrifuging briefly
6. Add 1ul of cDNA to the 384 well plate followed by 19ul of the master mix (plate layout below)
7. Seal the plate with an adhesive plate cover and centrifuge briefly
8. Load the plate onto the real time instrument

C. 384 Plate layout

D. Run the real-time PCR reaction on the 7900HT machine (Applied Biosystems) – FAST reaction

Stage	Temp (°C)	Time (mm:ss)
Hold	50	2:00
Hold	95	0:20
Cycle (40X)	95 60	0:01 0:20

E. Analyze the results

1. View the amplification plots for the entire plate
2. Omit any failures or C_T values within the triplicates that differ by more than 0.5 in value
3. Export the data and calculate the relative quantification of the samples relative to the controls and normalized to three housekeeping genes.

Gene name	ThermoFisher Assay number	Gene name	ThermoFisher Assay number
ALDOA	Hs00605198_g1	CXCR1	Hs01921207_s1
G6PD	Hs0016169_m1	CXCR2	Hs01891184_s1
DCXR	Hs00212433_m1	IGFBP-1	Hs00236877_m1
PGLS	Hs00359986_m1	IGFBP-2	Hs01040719_m1
TKT	Hs01115545_m1	IGFBP-6	Hs00181853_m1
C14ORF80	Hs00415039_m1	MMP12	Hs00159178_m1
AKT2	Hs01086099_m1	GLT1D1	Hs01087581_m1
GSK3A	Hs00997938_m1	GGT3P	Hs02387913_g1
KCNQ1	Hs00923522_m1	RPLPO	Hs00420895_gH
SLC25A22	Hs00368705_m1	ACTB	Hs01060665_g1
HPRT1	Hs02800695_m1		

APPENDIX G: EpiTech Methyl_II assay set up

Procedure

A. Restriction digestion

- Perform the restriction digestions using the EpiTect Methyl II DNA Restriction Kit (cat. no. 335452).
- Prepare a reaction mix without enzymes as indicated in Table 1 using 250 ng genomic DNA. 5x Restriction Digestion Buffer should be thawed and vortexed well before use

Table 1. Reaction mix without enzymes

Component	Volume (µl)
Genomic DNA (250ng)	3
5x Restriction Digestion Buffer	26
RNAse-free water	90
Final volume	120

- Add RNase-/DNase-free water to make the final volume 120 µl. Vortex to thoroughly mix the components and centrifuge briefly in a microcentrifuge.
- Set up 4 digestion reactions (M_o, M_s, M_d, and M_{sd}) according to Table 2.

IMPORTANT: All 4 tubes must contain equal amounts of genomic DNA.

Table 2. Restriction digestion

Component	Mo (µl)	Ms (µl)	Md (µl)	Msd (µl)
Reaction mix from step 3	28	28	28	28
Methylation-sensitive enzyme A	---	1	---	1
Methylation-dependent enzyme B	---	---	1	1
RNAse-free water	2	1	1	---
Final volume	**30**	**30**	**30**	**30**

- Pipet up and down to gently, but thoroughly mix the components. Centrifuge the tubes briefly in a microcentrifuge.
- Incubate all 4 tubes at 37°C for 6 h in a heating block or thermal cycler. The reaction can also be performed overnight.
- After incubation, stop the reactions by heat-inactivating the enzymes at 65°C for 20 min.
- Mix the samples thoroughly by vortexing before use. Centrifuge the samples briefly and proceed to step 1 of "Setting up the PCR".

B. Setting up the PCR

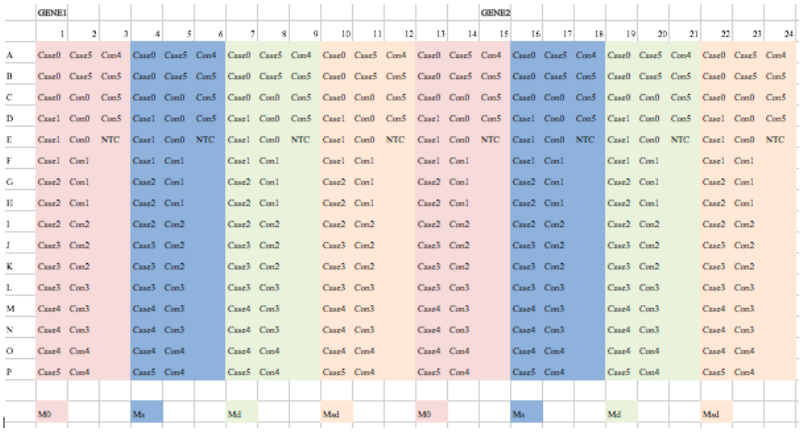
- Prepare individual reactions for each of the 4 digestions (M_o, M_s, M_d, and M_{sd}) in a 1.5 ml tube Table 3. Repeat for each gene.

Table 3: Setting up the PCR reaction

Component	Mo (µl)	Ms (µl)	Md (µl)	Msd (µl)
PCR Master mix	5	5	5	5
PCR primer mix	0.4	0.4	0.4	0.4
Mo digest	2	---	---	---
Ms digest	---	2	---	---
Md digest	---	---	2	---
Msd digest	---	---	---	2
RNAse-free water	2.6	2.6	2.6	2.6
Final volume	**10**	**10**	**10**	**10**

- Mix tubes well by vortexing, and briefly centrifuge the contents to the bottom of the tube.
- Add 25 µl of each reaction to the EpiTect Methyl II PCR Assay 384-well plate, as shown in Table 4.

Table 4. Assay setup (384 well plate)



- Seal or cap the wells of the plate. Centrifuge the plate for 1 min at 2000 rpm to remove any air bubbles.

C. Running the PCR

- Program the thermal cycler according to the manufacturer's instructions, using the conditions outlined in Table 5.

Table 5: PCR cycling conditions

Temperature (°C)	Time (mm:ss)	Number of cycles
95	10:00	1
99 72	00:30 1:00	3
97 72	00:15 1:00	40

Note: It is critical that the cycling conditions are followed exactly.

D. Data analysis

- **Obtaining raw threshold cycle (C_T) values**

After the cycling program has completed, obtain the C_T values according to the instructions provided by the manufacturer of the real-time PCR instrument. We recommend manually setting the baseline and threshold values as follows.

Baseline: Using the Linear View of the amplification plots, set the instrument to use the readings from cycle number 2 through the cycle just before the earliest visible amplification, usually between cycle 10 and 15.

Threshold value: Using the Log View of the amplification plots, place the threshold above the background signal but within the lower third of the linear portion of the amplification curves.

- **Exporting C_T values**

Export and/or copy/paste the C_T values from the instrument software to a blank Microsoft Excel spreadsheet according the manufacturer's instructions for the real-time PCR instrument.

- **Microsoft Excel based data analysis template**

Download the EpiTect Methyl II PCR Array Microsoft Excel based data analysis template, which is available at: www.sabiosciences.com/dna_methylation_data_analysis.php.

Then, paste in the C_T value data and analyze the automatically generated results by following the directions in the "Instructions" worksheet of the Excel file.

APPENDIX J: Resulting C_T values from the Epitech Methyl II assay

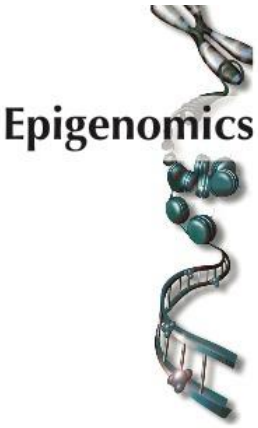
G6PD					TKT				
Blood samples	C_t mock	C_t dependent	C_t sensitive	C_t double	Blood samples	C_t mock	C_t dependent	C_t sensitive	C_t double
Case0	23.0292	24.0433	23.6153	33.5897	Case0	20.3782	25.0739	32.3435	35.1324
Case1	22.1956	22.7787	22.5777	28.1372	Case1	19.2912	19.7321	22.1520	26.3178
Case2	22.0966	22.4972	22.2787	28.3853	Case2	20.7312	22.7641	22.9049	25.7820
Case3	22.4648	24.9069	24.4452	29.1643	Case3	19.9659	20.8063	21.9747	27.7124
Case4	22.3247	33.3210	30.3704	35.0754	Case4	26.0000	38.0527	39.0826	38.4202
Case5	23.6415	33.1804	32.9852	33.2952	Case5	26.1251	30.0000	38.4896	36.4781
Control0	24.5362	33.3344	33.2832	34.3898	Control0	22.0145	35.0278	36.8130	38.2508
Control1	22.9918	23.9057	23.3242	33.7548	Control1	17.5101	20.8728	22.0315	28.3264
Control2	21.2924	23.5596	22.4729	25.6220	Control2	19.5590	21.0923	21.3141	29.4049
Control3	21.8398	22.9625	22.1087	32.7061	Control3	19.3429	25.3278	26.8379	30.9245
Control4	22.5291	35.2029	23.6402	35.5347	Control4	25.3424	31.0710	32.0930	34.8152
Control5	22.2884	23.1966	22.4044	26.8826	Control5	19.4256	36.7978	37.5080	37.6495
Placenta samples	C_t mock	C_t dependent	C_t sensitive	C_t double	Placenta samples	C_t mock	C_t dependent	C_t sensitive	C_t double
Case0	23.5185	30.3824	24.1697	33.8130	Case0	20.8103	26.8868	33.6914	34.9676
Case1	22.7333	28.5713	23.1625	34.2896	Case1	21.7623	22.1052	23.1132	28.3534
Case2	23.4623	24.9882	23.7616	29.5098	Case2	20.6610	26.5400	27.6126	28.8633
Case3	22.4648	24.9069	24.4452	29.1643	Case3	22.2972	22.4398	23.1741	32.2610
Case4	22.9744	23.6745	23.6765	33.7332	Case4	31.6102	34.6159	35.2930	38.7815
Case5	22.4099	34.9445	34.2646	35.7005	Case5	26.9592	38.8447	39.0826	36.6354
Control0	22.7333	28.5713	23.1625	34.2896	Control0	26.0184	34.2158	38.3675	36.7664
Control1	22.7333	28.5713	23.1625	34.2896	Control1	20.0548	24.7201	25.1752	30.7946
Control2	21.8445	28.7692	22.3988	34.1821	Control2	19.3336	19.8478	20.0709	27.3041
Control3	23.1127	31.0337	23.7487	34.3981	Control3	19.8822	25.8907	29.7100	34.4500
Control4	21.8367	31.2883	24.8550	34.2102	Control4	19.8598	25.2609	27.0382	36.5826
Control5	23.5841	33.5329	33.3436	33.5988	Control5	18.2360	35.0639	36.2930	39.5400
IGFBP-1					IGFBP-2				
Blood samples	C_t mock	C_t dependent	C_t sensitive	C_t double	Blood samples	C_t mock	C_t dependent	C_t sensitive	C_t double
Case0	20.6417	23.9191	25.6689	27.6247	Case0	26.7041	28.2084	31.0649	31.8627
Case1	19.7566	23.0952	26.2773	32.2685	Case1	19.7770	20.1458	21.3701	25.9992
Case2	19.7675	23.7134	25.3738	26.2900	Case2	18.5543	19.5624	20.4262	27.6179
Case3	20.2615	23.8999	27.3709	29.8063	Case3	20.4370	20.5647	21.1352	28.1358
Case4	17.9697	25.5424	27.2431	28.9455	Case4	27.8283	28.0829	28.2460	29.3471
Case5	19.5513	29.7438	30.7026	31.4737	Case5	27.4314	28.1975	28.8338	28.7243
Control0	24.2778	26.7620	29.5070	37.1891	Control0	21.6200	28.5066	28.8265	28.8388
Control1	21.2226	24.2624	25.5230	27.4402	Control1	23.1669	27.4318	27.8651	32.1460
Control2	19.1912	22.3788	24.9256	26.1503	Control2	19.6661	20.4275	21.1349	27.7121
Control3	19.6499	22.6375	26.0441	29.2666	Control3	19.3867	27.7932	27.9959	29.5384
Control4	20.7377	26.6046	28.9234	28.9913	Control4	27.6373	28.5171	28.5687	29.4155
Control5	20.4633	25.1450	28.2292	28.4545	Control5	19.4914	27.5329	27.8754	29.7569
Placenta samples	C_t mock	C_t dependent	C_t sensitive	C_t double	Placenta samples	C_t mock	C_t dependent	C_t sensitive	C_t double
Case0	19.9844	21.4869	21.8496	27.7701	Case0	21.5004	21.9148	21.9764	27.9325
Case1	21.3511	22.6944	24.3268	24.5046	Case1	20.7277	20.9449	21.7231	26.8116
Case2	20.2566	22.5040	22.6131	23.6095	Case2	20.7465	22.1768	22.2566	28.6351
Case3	21.6278	23.5710	25.6661	29.0029	Case3	26.9010	27.9419	28.2689	30.5213
Case4	22.5923	25.2584	25.5377	29.9465	Case4	27.4668	28.8268	28.9425	30.0289
Case5	20.5979	22.5487	24.0191	28.7715	Case5	19.8941	27.6681	27.6911	29.2139

Control0	24.0458	24.5533	29.4086	30.2717	**Control0**	26.0763	27.4079	27.8693	28.4594
Control1	17.5101	20.8728	22.0315	28.3264	**Control1**	20.0545	20.5441	20.8426	27.3665
Control2	19.8316	21.3249	22.1925	23.2579	**Control2**	20.5209	27.1854	27.5687	29.2093
Control3	19.9947	21.4134	22.8886	28.6740	**Control3**	20.7044	28.2878	28.4094	29.2762
Control4	21.6587	22.2174	24.2676	28.2867	**Control4**	21.0751	28.1524	28.7982	29.3413
Control5	19.1415	22.9440	28.3721	33.2693	**Control5**	20.7044	28.2878	28.4094	29.2762
IGFBP-6									
Blood samples	**C_t mock**	**C_t dependent**	**C_t sensitive**	**C_t double**					
Case0	20.3782	25.0739	32.3435	35.1324					
Case1	19.2912	19.7321	22.1520	26.3178					
Case2	20.7312	22.7641	22.9049	25.7820					
Case3	19.9659	20.8063	21.9747	27.7124					
Case4	26.0000	38.0527	39.0826	38.4202					
Case5	26.1251	29.0000	35.4896	36.4781					
Control0	22.0145	35.0278	36.8130	38.2508					
Control1	17.5101	20.8728	22.0315	28.3264					
Control2	19.5590	21.0923	21.3141	29.4049					
Control3	19.3429	25.3278	26.8379	30.9245					
Control4	25.3424	31.0710	32.0930	34.8152					
Control5	19.4256	36.7978	37.5080	37.6495					
Placenta samples	**C_t mock**	**C_t dependent**	**C_t sensitive**	**C_t double**					
Case0	20.8103	26.8868	33.6914	34.9676					
Case1	21.7623	22.1052	23.1132	28.3534					
Case2	20.6610	26.5400	27.6126	28.8633					
Case3	22.2972	22.4398	23.1741	32.2610					
Case4	26.6102	30.6159	31.2930	36.7815					
Case5	26.9592	35.8447	35.0826	36.6354					
Control0	32.0184	34.2158	38.3675	39.7664					
Control1	20.0548	24.7201	25.1752	30.7946					
Control2	19.3336	19.8478	20.0709	27.3041					
Control3	19.8822	25.8907	29.7100	34.4500					
Control4	19.8598	25.2609	27.0382	36.5826					
Control5	18.2360	35.0639	36.2930	39.5400					

APPENDIX K: Review article published in the journal of Epigenomics titled "Epigenetics and the burden of noncommunicable disease: a paucity for research in Africa"

Review

For reprint orders, please contact: reprints@futuremedicine.com



Epigenetics and the burden of noncommunicable disease: a paucity of research in Africa

Epidemiological evidence suggests that an adverse *in utero* environment is associated with an increased risk for developing adult onset diseases. The molecular mechanisms for susceptibility to chronic noncommunicable diseases are not fully understood, although recent research has proposed that epigenetic modifications play an important role in fetal programming. Genetic and environmental factors contribute to interindividual and spatiotemporal tissue-specific methylation patterns. Although the diverse environments and high genetic diversity of African populations provide unparalleled potential to investigate the effects of environmental change on the epigenetic profile in humans, only a small percentage of genomic and epigenetic studies have focused on populations from this continent. This emphasizes the need to build capacity in Africa for research that leads to an understanding of the association between genetic, epigenetic and environmental risk factors for noncommunicable diseases on the continent.

Keywords: Africa • DNA methylation • epigenetics • epigenome-wide association studies

Angela Hobbs[1]
& Michèle Ramsay*,[2]
[1]Division of Human Genetics, National Health Laboratory Service & School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]Division of Human Genetics, National Health Laboratory Service & School of Pathology, Faculty of Health Sciences & the Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa
*Author for correspondence:
Tel.: +27 11 717 6635
michele.ramsay@wits.ac.za

Noncommunicable disease in Africa

The burden of noncommunicable diseases (NCDs) is increasing globally and is currently the leading cause of death and disease burden worldwide. These NCDs not only pose a major challenge for healthcare systems but also to affected individuals [1]. Cardiovascular disease (CVD), cancer, chronic lung disease and diabetes result in more than 30 million deaths annually worldwide [1,2]. Previously regarded as diseases of high-income countries, NCDs are increasing at an alarming rate in low- and middle-income countries [3]. This is mainly due to demographic transitions and changing lifestyles of populations associated with urbanization. In Africa it is estimated that 40% of the population live in urban areas but by 2030, this number will exceed 50% as Africa ceases to be a predominantly rural continent. This transformation is evident by the high urban growth rate of 3.6% in sub-Saharan Africa, double the world average. This rapid urbanization has been accompanied by significant shifts in the health patterns of South Africans, thus increasing the prevalence of NCDs [4]. Although public health in African countries has focused primarily on communicable diseases such as HIV/AIDS, malaria and tuberculosis, NCDs are now major sources of morbidity and mortality and are projected to overtake infectious diseases by 2030 [1]. NCDs are to some extent preventable by modifying risk factors such as physical inactivity, high blood cholesterol, high blood pressure, obesity, unhealthy diet and inappropriate use of tobacco and alcohol [3,4]. These lifestyle related risk factors result in various long-term disease processes that lead to high mortality rates attributable to stroke, heart attacks, cancers, obstructive lung diseases and many others [4]. However, research also highlights the importance of *in utero* and early life environment to an increased susceptibility to developing NCDs later in life. Given the burden of NCD in many African countries where healthcare resources are scarce, it is important to identify pre-

Future Medicine part of fsg