Epigenetic mechanisms in the early life programming of obesity

James Casey McConnell

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Institute of Genetic Medicine

April 2012

Abstract

Obesity presents a major public health burden with prevalence rising in both children and adults. This disorder is associated with many adverse health outcomes and improved understanding of the mechanisms is required to develop effective preventive and treatment strategies.

It has been hypothesised that environmental exposures such as poor nutrition *in utero* and during the early post natal period can programme an individual to develop obesity in later life. These early life exposures can be 'memorised' by the cell in the form of epigenetic modifications, changes to the biochemical structure and function of DNA. Such modifications include DNA methylation, the addition of a methyl group to cytosine residues which is involved in the regulation of gene transcription. Epigenetic mechanisms therefore represent an attractive mechanism to explain developmental programming phenomena.

The overarching aim of this study was to establish the mediating role of epigenetic processes in linking modifiable environmental exposures with subsequent risk of obesity. This was addressed through interrogation of animal models, through the development and application of bioinformatic approaches and through epidemiological investigation of human population studies.

Tissue level DNA methylation patterns were investigated in hypothalamus using immunohistochemical staining. No significant differences were discernible between methylation levels in the hypothalami of control rodents when compared to hypothalami from rodents that had been exposed *in utero* to a dietary regimen that induces metabolic perturbation and obesity in offspring.

Bioinformatic approaches were used to develop and apply an *in silico* workflow to interrogate gene expression dataset, in this instance from a rodent model of dietary manipulation *in utero* and early postnatal life. The purpose of this *in silico* interrogation was to identify loci that were strong candidates for epigenetic regulation of gene expression. Four genes, *Esr1*, *Fxn*, *Igf2r* and *Rbl2* were identified and the levels of promoter methylation at these loci were assessed in rodent liver tissue from offspring of exposed and unexposed mothers using pyrosequencing. DNA methylation levels in *Igf2r* were observed to be higher in animals exposed to a maternal obesogenic diet.

Using epidemiological approaches, the relationship between obesity and related traits and DNA methylation was assessed. Firstly, DNA methylation levels in two candidate genes (*IGF2* and *TACSTD2*) believed to be associated with obesity were analysed in eight year old children. *IGF2* methylation was positively correlated with age, however neither locus demonstrated any association with body composition. Secondly, LINE-1 methylation, a surrogate for global DNA methylation, was assessed in a cohort of individuals aged 50 years. LINE-1 methylation was found to be associated with blood lipid and glycaemic markers including fasting glucose, total cholesterol, total triglycerides and LDL cholesterol and HDL:LDL ratio.

Early life predictors of DNA methylation were also explored in both childhood and adult cohorts. The relationship between DNA methylation in LINE-1 elements, *IGF2* and *TACSTD2* and gestational age, birth weight and length of breastfeeding were explored. Evidence of an association between gestational age at birth and DNA methylation patterns in both children and in adults in later life was observed.

Collectively these investigations provide some support for the hypothesis that epigenetic mechanisms, namely DNA methylation is a mediating mechanism linking environmental exposures during pregnancy and early life to the subsequent development of perturbed metabolic traits and possibly obesity. Further methodological developments to refine and expand the appraisal of DNA methylation patterns together with larger studies are required to extend these findings.

Contents

Index of figures

Index of tables

Abbreviations

5hMC	5-Hydroxymethylcytosine
5MC	5-Methylcytosine
AICDA/AID	Activation-induced cytidine deaminase
AD	Alzheimer's Disease
ARC	Arcuate nucleus, hypothalamus
BC	Bisulphite conversion
BIO	Biotinylated
BMI	Body mass index
BMR	Basal metabolic rate
BSMAP	Bisulphite sequence mapping program
BSPP	Bisulphite padlock probes
CART	Cocaine and amphetamine regulated transcript
CGI	CpG island
CI	Confidence interval
CNS	Central nervous system
COPD	Chronic obstructive pulmonary disease
CVD	Cardiovascular disease
CyMATE	Cytosine methylation analysis tool for everyone
DAB	3,3'-Diaminobenzidine
DAVID	Database for Annotation, Visualization and Integrated Discovery
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DPX	Distyrene plasticizer xylene
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
EMBOSS	European molecular biology open software suite

ER	Endoplasmic reticulum
FDR	False discovery rate
FITC	Fluorescein isothiocyanate
FRDA	Friedreich's ataxia
FTO	Fat mass and obesity-associated protein
FU	Fluorescence unit
FXN	Frataxin
GDLD	Gelatinous drop-like corneal dystrophy
GMS	Gateshead Millennium Study
GO	Gene ontology
GWAS	Genome-wide association study
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDL	High density lipoprotein
HOMA	Homeostasis model assessment
HPRD	Human protein reference database
IGF	Insulin-like growth factor
IGFIIR	Insulin-like growth factor 2 receptor
IHC	Immunohistochemistry
IHD	Ischaemic heart disease
IHTG	Intrahepatic triglyceride
IPA	Ingenuity pathway analysis
IQR	Interquartile range
IR	Insulin receptor
IVM	In vitro methylated
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LDL	Low density lipoprotein
LINE	Long interspersed nuclear element

LTR	Long terminal repeat
MBD	Methyl binding domain
MeCP2	Methyl CpG binding protein 2
MeDIP	Methylated DNA immunoprecipitation
MEDME	Modelling experimental data with MeDIP enrichment
MIB	Monoclonal antibody to Ki-67
MIT	MeInfoText
MONICA	Multinational monitoring of trends and determinants in cardiovascular disease
NAFLD	Non-alcoholic fatty liver disease
NCBI	National Centre for Biotechnology Information
NHS	National Health Service
NIH	US National Institutes of Health;
NPA	No primary antibody
NPY	Neuropeptide Y
NSE	Neurone specific enolase
NTFS	Newcastle Thousand Families Study
OMIM	Online Mendelian Inheritance in Man
PAF	RNA polymerase II-associated factor
PAH	Polycyclic aromatic hydrocarbons
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFOS	Perfluorooctane sulfonate
PLP	Postnatal low protein
POMC	Pro-opiomelanocortin
PPAR	Peroxisome proliferator-activated receptor
PROBIT	Promotion of Breastfeeding Intervention Trial
PSQ	Pyrosequencing software for quantification

QUMA	Quantification tool for methylation analysis
RB	Retinoblastoma
RCT	Randomised control trial
RGB	Red green blue
RNA	Ribose nucleic acid
RRBS	Reduced representation bisulphite sequencing
SAH	S-Adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SD	Standard deviation
SNP	Single nucleotide polymorphism
TACSTD	Tumour-associated calcium signal transducer
TAQ	*Thermus aquaticus*
TCGA	The cancer genome atlas
TDM	Text data mining
TI	Thrive index
TSS	Transcription start site
UK	United Kingdom
US	United States
VBA	Visual basic analysis
VPW	Vacuum prep workstation
WAT	White adipose tissue
WGA	Whole genome amplification
WGSBS	Whole-genome shotgun bisulphite sequencing
WHO	World Health Organisation

Acknowledgements

There are a number of people without whose help this thesis would not have been possible.

Alix, Caroline, Hannah, John, Julie, Kate, Laura. Thanks guys.

Chapter 1: Introduction

1.1 Obesity

1.1.1 Definition

Overweight and obesity are defined as abnormal or excessive fat accumulation that may impair health (WHO, 2011). Both conditions are commonly classified using Body Mass Index (BMI), a simple index of weight-for-height, defined as weight in kilograms divided by height in metres squared (kg/m^2) (Dietz and Robinson, 1998). A normal BMI is 18.5-24.5 kg/m^2 (Cole et al., 2000), with adult humans classified as being overweight if BMI is $25kg/m^2$ or above and clinically obese if BMI is $30kg/m^2$ or above (Visscher et al., 2010). Although BMI is widely used as a measure of adiposity there is some evidence that is may not be the best way to assess weight gain. BMI has a number of limitations in that it is only a proxy indicator of body fatness and factors such as fitness, ethnic origin and pubertal status can alter the relationship between BMI and adiposity. BMI also does not fully adjust for height or body shape. As such some studies use other measures such as waist or hip circumference, body fat ratio and skin fold thickness, however these measures are more difficult and expensive to collect in large numbers and as such BMI has become the established routine measure of adiposity. Obesity which is characterised by an increase in central fat mass is associated with a number of comorbidities including chronically increased blood pressure and the perturbation of circulating levels of both lipids and glucose (Ginsberg and MacCallum, 2009).

1.1.2 Epidemiology

Worldwide prevalence of overweight and obesity has been rising steadily for the last 40 years in both children (Figure 1.1) (Swinburn et al., 2011) and adults (Figure 1.2) (Sassi, 2010). This upsurge is largely attributable to an increasingly Westernised diet with elevated availability of high fat, high sugar foods coupled with an increasingly sedentary lifestyle. Consequently obesity has become one of the most serious health concerns of our time. Globally obesity is now the fifth leading risk of death and at least 2.8 million adults die each year as a result of being overweight or obese (WHO, 2011). The obesity epidemic is observed not only in adults but also in children in both the developed and developing world (Gupta et al., 2012). Indeed overweight and obesity are now linked to more deaths worldwide than underweight.

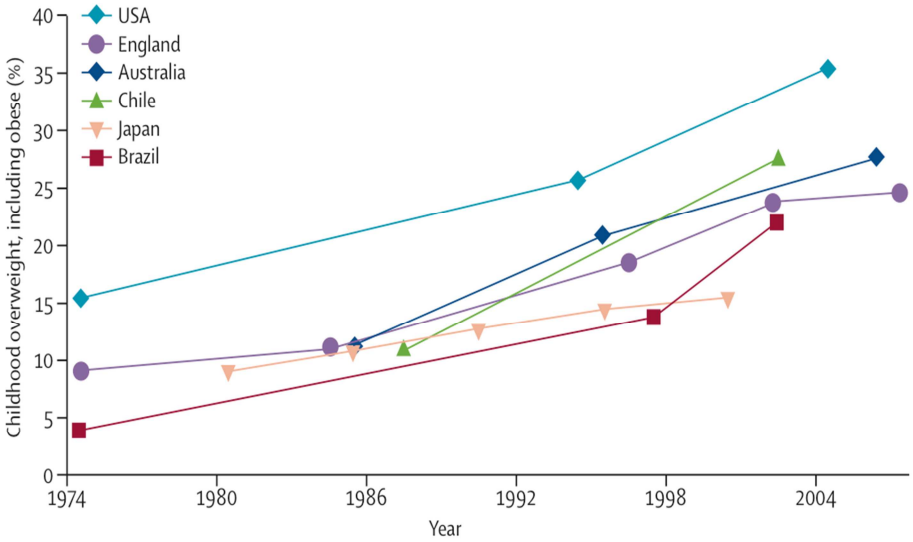


Figure 1.1. Worldwide estimates of percentage childhood overweight and obesity. Using International Obesity Taskforce cut offs (Swinburn et al., 2011).

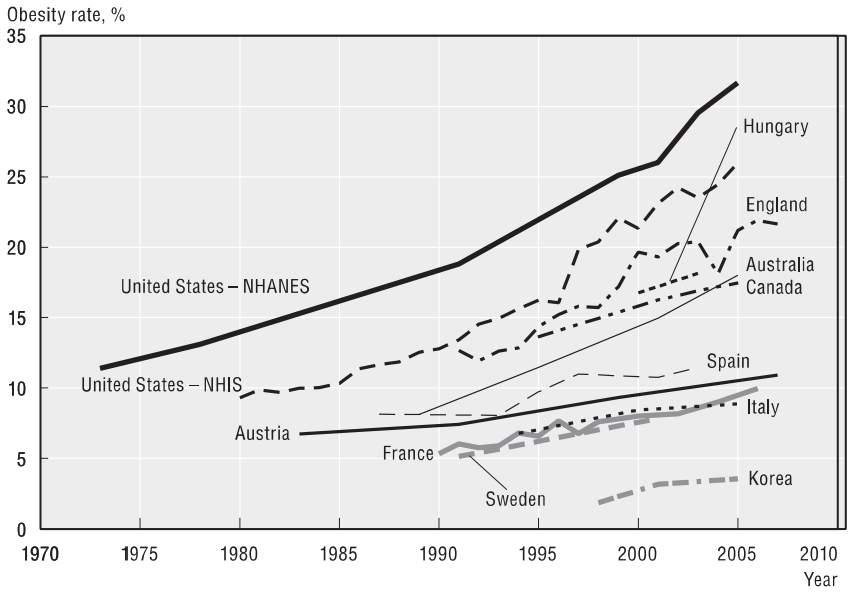


Figure 1.2. Age-standardised obesity rates, age 15-64, selected countries. (Sassi, 2010)

1.1.3 Prevention and treatment of obesity

The health burden of obesity has led to attempts to lower its prevalence at a population level (Miller et al., 2002). Energy balance plays a fundamental role in the development of obesity and, when energy intake exceeds energy expenditure over a sustained period, this positive energy balance can result in an obese phenotype. Consequently, recommended interventions to decrease the likelihood of developing obesity include diet modification to reduce the level of fat and sugar in an attempt to decrease daily energy intake and therefore promote weight maintenance or loss rather than weight gain (Ross et al., 2000). Another possible intervention to prevent the onset of an obese phenotype is augmented levels of mobility and exercise. Participation in more exercise results in increased energy expenditure, again encouraging weight loss (Shaw et al., 2006). Individuals with severe obesity (BMI >40kg/m^2) may find it difficult to sustain the behavioural changes of diet and exercise modification required to reduce their body weight. In these circumstances bariatric surgery such as gastric banding may be offered (Buchwald et al., 2004), whereby the stomach is banded or stapled to reduce its size. Bariatric surgery acts to physically reduce the capacity of the digestive tract, limiting either an individual's capacity for food intake (Santry et al., 2005) or their ability to digest and absorb dietary components.

1.1.4 Associated health problems

The current prevalence of obesity places a heavy burden on health services because obesity increases the risk of development of a number of conditions including type 2 diabetes and cancer which are becoming more prevalent in contemporary society (Berenson, 2012, Allender and Rayner, 2011). It is estimated that 44% of the diabetes burden, 23% of the ischaemic heart disease burden and between 7% and 41% of certain cancer burdens are attributable to overweight and obesity (WHO, 2011).

1.1.5 The development of obesity

Environmental factors such as a diet high in fats and sugars and lack of exercise are instrumental in the development of overweight and obesity (Utter et al., 1998). Worldwide the average amount of food purchased per person per day has risen year on year over the last 40 years. There has been a worldwide shift toward a more sedentary lifestyle as people do less physically demanding jobs and take less exercise due to greater use of mechanised transport (Martinez-Gonzalez et al., 1999). Although environmental factors such as diet and exercise play a key role in the development of obesity, genetic factors also exert an effect (Farooqi and O'Rahilly, 2007). Single nucleotide polymorphisms (SNPs) in genes implicated in the control of appetite and metabolic rate identified in a number of recent genome wide association studies (GWAS) can also predispose an individual to obesity (Li et al., 2010a).

Clearly, an individual's risk of developing obesity and related sequelae can be influenced by early life exposures. Evidence from studies in animal models demonstrates that obesity and its metabolic sequelae are 'programmed' during the prenatal or early postnatal period as a result of *in utero* exposures such as maternal under-nutrition (Martin-Gronert and Ozanne, 2010). These finding are supported by humans studies which show that the altered nutrition *in utero* results in offspring more likely to develop an obese or metabolic disease phenotype in later life (Ravelli et al., 1976, Barker et al., 1993).

1.1.6 A molecular basis

Obesity can be succinctly described as pathology of energy homeostasis (Spiegelman and Flier, 2001). Individuals who express phenotypic changes characteristic of obesity must undergo increased energy intake, decreased energy expenditure or indeed a combination of the two over a sustained period of time, leading to long-term positive energy balance. An individual's energy expenditure is determined both by their basal metabolic rate (BMR) and their level of physical activity (Lazzer et al., 2009). BMR is defined as the rate at which an organism uses energy when at complete rest, which in humans is measured by the heat given off per unit time, and expressed as the calories released per kilogram of body weight per hour (Black, 2000). Chronically low levels of physical activity result in decreased energy expenditure and as a result are likely to be obesogenic. Several biological pathways are likely to be involved in this process. These include the perturbation of appetite regulation, regulated by altered sensitivity to leptin, aberrant energy utilisation, controlled by perturbed

glucose-insulin signalling and the deposition of fat tissue caused by anomalous adipocyte metabolism. The multiple mechanisms at play are all likely to contribute to the pathogenesis of obesity to some degree (Figure 1.3) (Vickers and Sloboda, 2010). The organs of the body implicated in the development of obesity are also shown in Figure 1.3 and dealt with in detail in the following section of this chapter.

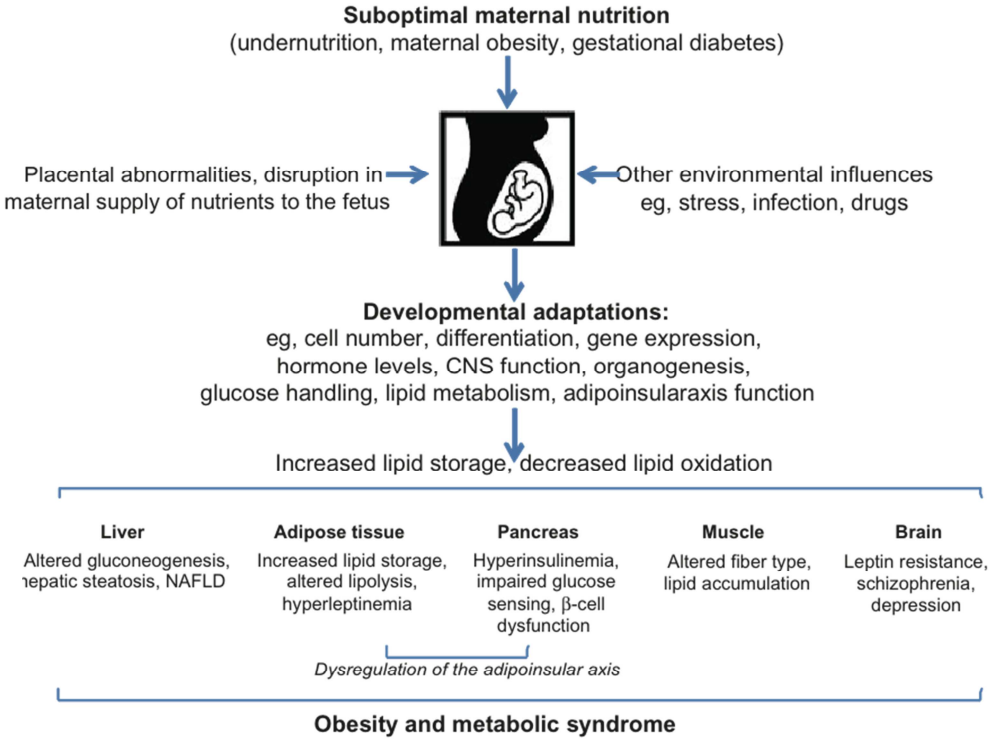


Figure 1.3 Pathways to obesity and related conditions following early life insult.
CNS, central nervous system, NAFLD, non-alcoholic fatty liver disease. Adapted from (Vickers and Sloboda, 2010).

1.1.6.1 Liver

The liver is involved in a whole host of biological processes. It is involved in the pathogenic processes associated with obesity including the development of hepatic steatosis and therefore non-alcoholic fatty liver disease (NAFLD), a pathology closely linked with aberrant gluconeogenesis and fatty acid metabolism (Bosserhoff and Hellerbrand, 2011). Steatosis is a key characteristic of NAFLD and occurs when the rate of fatty acid uptake by hepatocytes is greater than the rate of fatty acid oxidation and export from the liver, leading to increased accumulation of intrahepatic triglyceride (IHTG) (Ndumele et al., 2011). Steatosis in liver tissue has been associated with altered glucose, fatty acid, and lipoprotein metabolism (Chan et al., 2010). Abnormalities in these key metabolic pathways form important factors in the pathogenesis of insulin resistance,

dyslipidaemia, and other cardiometabolic risk factors associated with obesity (Chatrath et al., 2012). The temporality of these changes in the liver are not well characterised however, in as much as it is not clear if NAFLD is the cause of metabolic dysfunction or if metabolic dysfunction results in increased IHTG accumulation.

1.1.6.2 Adipose tissue

Leptin is a hormone secreted in the most part by white adipose tissue (WAT), with high levels of plasma leptin strongly associated with the obese phenotype (Aguilera et al., 2008). The leptin signalling cascade can act to reduce the synthesis of free fatty acids, preventing their deposition as new WAT (Scherer and Buettner, 2011) and as such can result in decreased adipose tissue lay-down (Buettner et al., 2008). Not only does leptin help control fat deposition, it also binds leptin receptors in the arcurate nucleus , leading to inhibition of neuropeptide Y (NPY) neurons (Wang et al., 1997), acting to increase satiety and therefore resulting in reduced dietary energy intake (Leinninger et al., 2009). Leptin's key role as both a regulator of adiposity and appetite mean that a perturbation of the leptin-signalling pathway (Considine, 2011) leads to dysregulation of adipogenesis and/or chronically increased appetite levels, leading to the development of obesity (Montague et al., 1997).

A further pathway implicated in the development of obesity is altered adipocyte metabolism. The function of adipocytes in the context of WAT as an endocrine organ is not yet fully understood, however their role in the storage of triglycerides and fatty acids is well characterised (Trayhurn and Beattie, 2001, Wood et al., 2009). Adipose tissue is comprised mainly of adipocytes, the population of which is dependent on the body's fat storage needs (Ruge et al., 2009). Under normal conditions adipocytes undergo hyperplasia and hypertrophy in response to increased circulating fatty acid concentrations and provide greater lipid storage capacity.

A loss of function in genes controlling adipocyte proliferation gives rise to decreased adipocyte hyperplasia. This increase in adipocyte number and size has a number of consequences. Firstly, as adipocytes are endocrine cells which secrete leptin, an increased cell mass will result in the disruption of the leptin signalling pathway, causing downstream consequences in both appetite regulation and control of adiposity. Secondly, increased adipocyte size and number will result in a subsequent deposition of fat in non-adipose tissues such as skeletal muscle, pancreas and liver (Reitman et al., 1999). The deposition of non-adipose tissue fat has been strongly associated with the development of obesity-related disorders including type 2 diabetes (Cali and Caprio, 2009).

1.1.6.3 Pancreas

Insulin is a key hormone in the control of circulating blood glucose levels (DeFronzo and Ferrannini, 1991). Released by the pancreas in response to raised concentrations of plasma glucose (above fasting levels of approximately 5mM), insulin results in a decrease in gluconeogenesis in the liver and an increase in glucose uptake in muscle. Forming a feedback loop alongside glucagon, insulin allows blood glucose levels to be controlled homeostatically (Matthews et al., 1985). An individual's glucose tolerance is determined by both their insulin secretion and insulin sensitivity. Low insulin sensitivity, also termed insulin resistance, results in the dysregulation of blood glucose levels and increases the risk of developing a pre-diabetic or diabetic phenotype (Haffner et al., 1992). Impaired interactions between glucose and the insulin signalling pathway have been strongly associated with the obese phenotype (Sinha et al., 2002) and as such it has been heavily studied in the context of early life exposures, particularly nutritional insults. One study showed that in an older cohort (mean age 69.5 years) insulin resistance, as measured by oral glucose tolerance test, was associated with low birth weight, thinness at birth and low maternal BMI. This impaired insulin resistance was also shown to be associated with an increased risk of overweight and obesity in later life (Eriksson et al., 2002). A possible explanation for this adverse effect of increased weight gain is that foetal growth restriction may lead to reduced pancreatic cell growth (Blondeau et al., 2002). The result of this may be the development of a pancreas unable to respond to the metabolic demand of accelerated growth during childhood, predisposing the individual to problems with glucose homeostasis.

1.1.6.4 Skeletal muscle

The musculature has been implicated in the development of obesity as it has been shown that obesity causes muscle fibre changes as well as altered glucose utilisation (Mitrou et al., 2009). Skeletal muscle is the principal site of fatty acid and glucose metabolism in the body. Very flexible, it is able to switch from mainly lipid oxidation during conditions of fast to increased levels of glucose uptake, storage and oxidation when stimulated with insulin. Muscle fibres which make up the skeletal muscle fall into three main categories, Types I, IIa and IIb. Type I fibres are known as 'slow twitch' fibres. This subset contains high numbers of mitochondria and are therefore oxidative.

Type IIa fibres are 'fast twitch' and are high in both mitochondrial and glycogenolytic enzymes, whereas IIb are high only in glycogenolytic enzymes (Janovska et al., 2010).

Oxidative muscle fibres such as Type I fibres predominantly use lipids as their substrate, whereas Type IIa fibres use glucose, and the IIb a mixture of the two (Coen et al., 2010). Muscles containing high levels of Type I fibres are more sensitive to insulin and consequently tend to contain higher levels of triglycerides (He et al., 2001). Obesity and type 2 diabetes have been associated with a decrease in Type I muscle fibre levels and an increase in Type IIb muscle fibre levels (Nyholm et al., 1997). In these circumstances skeletal muscle is less able to metabolise its substrates, leading to alterations in circulating levels of glucose (Ritov et al., 2010) and lipids (Eckardt et al., 2011).

1.1.6.5 Brain

As previously mentioned, leptin resistance and obesity have been closely linked, however the causal pathway is controversial (Myers et al., 2010). What is clear, however, is that the brain is involved in the regulation of leptin resistance through the action of Proopiomelanocortin (*POMC)* neurones (Vong et al., 2011) as well as being involved in more subtle processes such as the development of mental and psychological disorders which lead to altered feeding habits (Berridge et al., 2010). There are a number of hormonal and neural mechanisms through which the brain is able to sense levels of ingested and stored nutrients, reacting to them by controlling behavioural, autonomic, and endocrine output. The hypothalamus and caudal brainstem play crucial roles in homeostatic function (Blevins and Baskin, 2010), acting to moderate hunger via the leptin pathway as previously discussed. The cortex and limbic system perform a more nuanced role, processing information on memory, reward and emotion, as well as placing food in a social and environmental context (Epel et al., 2011). Alterations in any of these complex hormonal and autonomic pathways can result in excess energy intake and increased fat storage (Bermudez-Silva et al., 2012). This process is one of a number of pathways that may elicit the development of obesity and related sequelae. Subsequently, a more thorough understanding of the role the brain plays will help target intervention strategies to reduce the prevalence of obesity.

1.1.7 Altered gene expression and obesity

One way in which early life exposures may affect the above molecular mechanisms is by altering the expression levels of genes within the pathways. Up or down regulation of a key gene may result in the pathway function being impaired. Studies using this approach utilise microarray technology in order to assess expression levels of putative target genes. One example showed differential expression of genes involved in adipogenesis, including

Peroxisome proliferator-activated receptor gamma (PPAR-γ) in subcutaneous adipose tissue of morbidly obese humans (Rodriguez-Acebes et al., 2010). A wide range of genes, some of which are reviewed below, has been shown to exhibit differential expression and be associated with the development of obesity.

1.1.7.1 *FTO*

The Fat mass and obesity related gene (*FTO*) has been unequivocally linked to the development of obesity and related conditions: Common single nucleotide polymorphisms (SNPs) in the first exon of *FTO* are associated with measures of obesity including BMI, hip circumference, and total body weight in humans (Scuteri et al., 2007), however the molecular function of *FTO* has not been well established. A recent study has found that mice showing decreased expression of *FTO* show a significant reduction in adipose tissue and lean body mass. These mice were also shown to exhibit increased energy expenditure (Fischer et al., 2009). Similar findings have been reported in humans, where *FTO* gene expression was negatively correlated with measures of obesity including BMI and percentage body fat (Kloting et al., 2008). Studies have also shown that *FTO* exhibits demethylase activity and that expression of the *FTO* risk allele is associated with genome wide DNA methylation changes (Almen et al., 2012) Taken together, these findings suggest that *FTO* is involved in energy homeostasis through the control of energy expenditure and may be acting via epigenetic mechanisms.

1.1.7.2 *POMC*

POMC, another gene in which both SNPs (Challis et al., 2002) and differential gene expression have been associated with the development of obesity, is expressed in the hypothalamus (Gee et al., 1983) and is a direct target for leptin (Cheung et al., 1997), as discussed earlier. As such it is implicated in the control of appetite (Pritchard et al., 2002). Studies have shown that mice not expressing *POMC* are obese and hyperphagic and are not able to react to increased fat content in food, resulting in weight gain (Challis et al., 2004). More recent work in rats has shown that *POMC* mRNA expression increased in response to pre-natal under-nutrition and postnatal high-fat nutrition (Ikenasio-Thorpe et al., 2007). These studies suggest that differential gene expression at the *POMC* locus might play a role in the pathogenesis of obesity through altered appetite regulation.

1.1.7.3 *PPAR-γ*

PPAR-γ is a regulator of adipocyte differentiation (Rosen et al., 1999). Intimately involved in the control of body fat deposition (Kubota et al., 1999), studies in mice show that knocking down *PPAR-γ* results in the abolition of adipose tissue synthesis in response to a high fat diet (Jones et al., 2005). In human studies *PPAR-γ* mRNA expression in skeletal muscle has been shown to be elevated in response to both increased BMI and increased fasting insulin levels (Park et al., 1997). This suggests that *PPAR-γ* may be involved in the development of insulin resistance and therefore type 2 diabetes and obesity.

1.1.7.4 Investigating differential gene expression in obesity

There are a plethora of differentially expressed genes implicated it the pathogenesis of obesity (van Dijk et al., 2009, Catalan et al., 2011). Interrogating these candidates may help improve our understanding of the mechanisms underlying this process and assist in the development of interventions. The temporality of changes is also key, with early life exposures having been shown to be very clearly associated with an individual's risk of developing obesity and related sequelae (Gluckman et al., 2008). The implications of these findings are discussed in the following section.

1.2 Developmental programming of health and disease – the evidence

1.2.1 Maternal nutrition

As previously discussed, obesity has reached epidemic proportions in human populations, an increase explained in part by over-nutrition and lack of exercise. Strikingly, this trend is borne out in pregnant women with marked increases in the BMI of women worldwide entering pregnancy, evidenced in one recent study (Figure 1.4) (Simmons, 2011). This rise in maternal pre-pregnancy weight presents an obvious risk to the health of the mother, increasing risk of a range of pathologies. These include and increased risk of cardiovascular disease and stroke (Poirier et al., 2006). It is also increases the likelihood of an adverse outcome during the pregnancy, such as the development of gestational diabetes (Chu et al., 2007), pre-ecalmpsia (Walsh, 2007), preterm labour and nephropathy. There are however other more persistent problems which may be caused by maternal obesity during pregnancy. There is increasing evidence from animal studies that excess maternal nutrition can programme the foetus to develop cardiovascular disease and other conditions in later life (Boney et al., 2005).

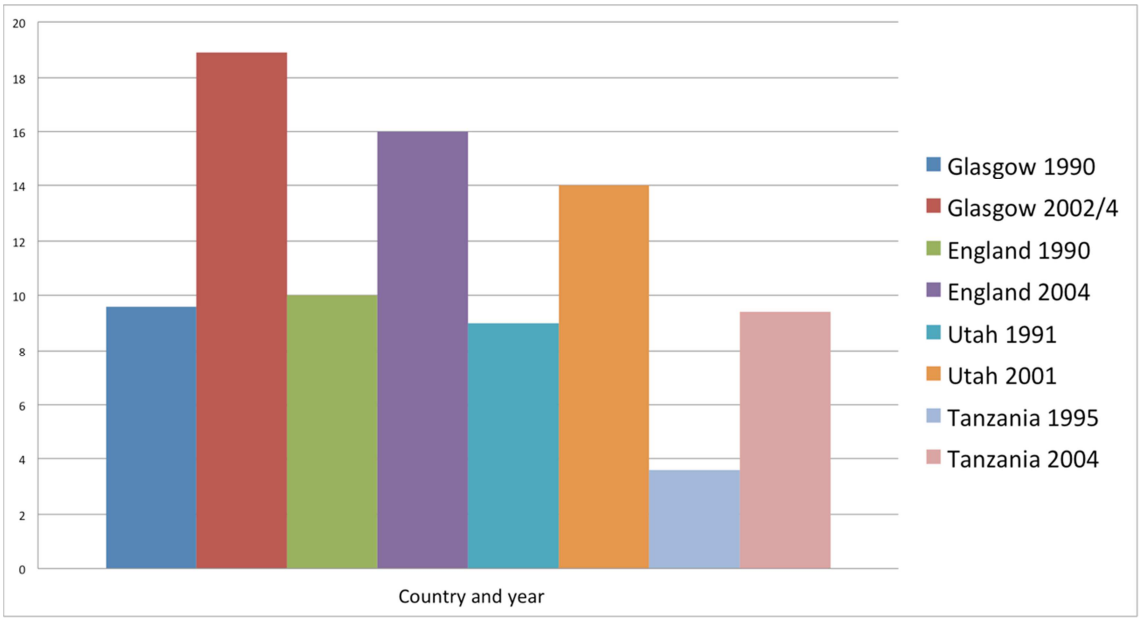


Figure 1.4 Prevalence of maternal obesity.
Four different international study centres are shown comparing percentage of obese women attending (BMI >30kg/m^2) during pregnancy across a range of years (Simmons, 2011).

1.2.2 Early theories in developmental programming

The effect of maternal nutrition on the foetus and its subsequent long-term health has been well characterised in both population based studies and animal experimental models. Under-nutrition has been associated with low birth weight, catch up growth and the subsequent development of the metabolic syndrome as an adult (Fagerberg et al., 2004). An early hypothesis proposed that nutritional insult *in utero* leads to permanent changes in insulin signalling and glucose metabolism, increasing the risk of later development of type 2 diabetes (Hales and Barker, 1992). This concept, known as the 'Thrifty Phenotype' hypothesis, is supported by studies showing that maternal nutrition restriction increases the risk of developing a range of pathologies including obesity (Ravelli et al., 1999). An extension of the 'Thrifty Phenotype' hypothesis came in the form of the 'Predictive Adaptive Response' hypothesis (Gluckman and Hanson, 2004a). This proposes that the offspring makes adaptions *in utero* or in in the early postnatal period in response to nutritional exposure. In this way, the foetus is 'predicting' what its nutritional exposures will be during early life. If the foetus is exposed to poor or excess nutrition *in utero* and this exposure is borne out in later life, then the phenotype will be normal. If however the exposure changes from high to low nutrition or vice versa then this is termed a mismatch and a disease phenotype may be manifest.

1.2.3 The DOHaD hypothesis

The Developmental Origins of Health and Disease hypothesis (DOHaD), synthesised from the early literature on the foetal origins of later health outcomes, proposes that early life exposures predispose an individual to the development of disease in later life (Gluckman and Hanson, 2006). A key observation brought about by the application of this hypothesis is that insult *in utero* and during lactation is associated with an increased risk of cardiovascular disease in later life (Hales and Barker, 2001). This observation has since been extended to include a variety of conditions including cognitive impairment (Van Den Bergh, 2011) and kidney disease (Vehaskari, 2010) although discussion here will be limited to the development of obesity, this being the focus of this thesis.

This hypothesis is relevant to the development of obesity, as it has been shown that obesity risk can be inferred from birth weight (Cnattingius et al., 2011). What is interesting to note is that the relationship is not linear but rather U shaped; the greatest risk for the development of obesity lies at the upper and lower ends of the birth weight range (Parsons

et al., 1999). Furthermore it has been shown that this birth weight effect can be exacerbated if there is significant disparity between an individual's nutritional environment during gestation and in later life –the mismatch effect alluded to above (Hanson and Gluckman, 2008). These findings are tempered by a contrasting body of literature which suggests that it is the overall cumulative exposure during the lifecourse rather than early life exposures which predispose an individual to disease in later life.

What is clear, and has become embodied in contemporary research into common complex diseases, is that changes in nutritional status during early life have profound and far-reaching effects on health in later life, including obesity. A more thorough understanding of the mechanisms through which this occurs in relation to the development of obesity could have significant public health benefits. There is evidence in both human and animal studies that early life exposures predispose an individual to obesity and related conditions such as type 2 diabetes and this is summarised in the following sections.

1.2.4 Human Studies

Studies into the effects of developmental programming in humans are complicated by a wide variety of factors including confounding by an individual's genotype and by environmental exposures during the life course. A number of strategies for dealing with these issues have been developed, and these are discussed in the following sections.

1.2.4.1 Studies of *in utero* exposure

Studying the effects of maternal nutrition and therefore *in utero* exposure in humans has proved challenging, not least because implementing dietary restriction on pregnant mothers is potentially hazardous to the health of both mother and child and therefore unlikely to meet stringent ethical standards. Consequently the majority of studies into maternal nutrition utilise cohorts in which offspring were *in utero* during conditions of dietary insult, such as famine. One cohort used extensively in these studies, known as the Dutch Hunger Winter cohort, comprises some 40,000 children conceived and born in 1944-45 who were exposed to famine *in utero* as a result of wartime food blockade (Stein et al., 1975). This cohort has proved useful in the assessment of the effects of early life nutrition on the development of disease phenotypes, with the first study to link nutritional deficit in the programming of obesity being performed on this group (Ravelli et al., 1976). A later study on a different subset of this cohort found that, in individuals aged 50 years, those that were

in utero during the five month famine period exhibited lower levels of glucose tolerance than individuals born either before or after the famine (Ravelli et al., 1998).

Large-scale famines of this type are relatively rare and as such findings in one study can prove difficult to replicate in other studies. For example, in the Finnish famine cohort, a study of nearly 900,000 individuals born around the time of the severe 1866–1868 famine, individuals experiencing extreme nutritional deprivation *in utero* experience a rise in mortality in early life but went on to have a lifespan not significantly different from non-exposed individuals (Kannisto et al., 1997). However, a cohort of individuals subjected to the 'Great Chinese famine' of 1959-61 has also been used to investigate the effects of suboptimal early life nutrition on later health. One study in this cohort showed that exposure to famine *in utero* leads to increased risk of hyperglycaemia (Li et al., 2010b), while a more recent study of 7,874 adults born between 1954 and 1964 in this cohort showed that exposure to famine in foetal life or infancy was associated with an increased risk of metabolic syndrome in adulthood (Li et al., 2011). Studies in famine cohorts have demonstrated that early life exposure to under-nutrition results in the development of a type 2 diabetes/metabolic disease phenotype. This begs the question of whether it is the nutritional insult itself or the mismatch with normal nutrition in later life that causes the onset of symptoms. If it is the latter then these studies add further credibility to the predictive adaptive response theory outlined above (Gluckman and Hanson, 2004b).

1.2.4.2 Studies of early postnatal life

Although in humans the *in utero* period is seen as key in the programming of disease susceptibility in later life, there is increasing evidence that the early postnatal period also exerts an effect. The first tentative observations along these lines came from the previously mentioned Dutch Hunger Winter cohort, where reduced nutrient intake during the first months of life significantly reduced obesity risk at age 19 years (Ravelli et al., 1976). However, like *in utero* exposure in humans, early life nutritional exposures can prove difficult to study. Consequently many studies are observational and often assess the effect of established exposures such as breast-feeding on the risk of obesity in later life. One such study found that in a cohort of some 9357 children aged five and 6, the prevalence of obesity in individuals who had never been breast fed was 4.5% compared to 2.8% in breastfed children. A dose-response effect was also identified for the duration of breast feeding on the prevalence of obesity with a prevalence of 3.8% for two months of breast feeding, 2.3% for 3-5 months, 1.7% for 6-12 months, and 0.8% for greater than 12 months (Von Kries et al., 1999). The findings of this initial study were backed up by further work

by the same group which found in a systematic review that breast feeding exerts a small but consistent protective effect against the development of obesity (Arenz et al., 2004).

Interesting though the findings of these observational studies are, their design means they are unable to show causality. A recent study comparing a number of cohorts with differing confounding structures showed, for example, that breast feeding has no effect on obesity risk (Brion et al., 2011). Randomised control trials (RCTs) however do allow causation to be inferred. The PROBIT (Promotion of breastfeeding intervention) trial of randomisation to breast feeding, the largest of its kind to date (Tilling et al., 2011), also indicates that breastfeeding is not associated with childhood obesity, although these data are yet to be published. An example of an RCT trial showing the effect of early postnatal nutrition on obesity risk is a recent study which showed that in two separate RCTs, exposure of infants to a protein and energy enriched diet increased fat mass later in childhood (Singhal et al., 2010). This study coupled with the observational data mentioned above support a causal link between over-nutrition and faster weight gain in early postnatal life and a greater risk of obesity and related sequelae in later life.

1.2.4.3 Twin studies

Twin studies are well established as a method of comparing environmental exposures in two genetically identical individuals (Naukkarinen et al., 2011). They have provided strong evidence to support the role of early life exposures in the developmental programming of obesity related conditions. One such study, in twins aged 55-74 years, showed that in both monozygotic and dizygotic twins of whom only a single twin had type 2 diabetes, the diabetic twin had a lower than average birth weight than their non-affected sibling (Poulsen et al., 1997). These findings have been replicated in younger twins (mean age 32.4 years) (Bo et al., 2000). The reason twin studies are ideal for this kind of assessment is that, accepting monozygotic twins are genetically identical (Machin, 2009), the differences in birth weight between twin pairs can be attributed to exposures such as access to nutrients *in utero* or postnatally. Similarly, a study of monozygotic female twins aged 18-34 years showed a negative correlation between birth weight and measures of body composition including waist/hip ratio and skin fold thickness when twins were compared (Loos et al., 2001). Again, this suggests that birth weight plays a role in programming of body composition in later life. Taken together these studies provide strong evidence for the importance of the foetal environment in mediating the relationship between birth weight and later development of diabetes.

1.2.5 Animal Models

There is burgeoning evidence from animal models in support of the DOHaD hypothesis (McMullen and Mostyn, 2009), due to the fact that they represent an ideal opportunity to investigate exposures during the lifecourse, particularly during the critical windows of gestation and early life, and later phenotypic manifestations.

1.2.5.1 Under-nutrition *in utero*

Nutritional insult *in utero* has been heavily studied in animal models, particularly in mice and rats. A large number of studies have used dietary restriction as model or under-nutrition *in utero*. An early study in rats showed that offspring of dams fed a 50% nutrient restricted diet gained more weight and showed adipocyte hypertrophy than control animals when weaned onto a normal diet (Jones et al., 1984). This represents another example of the mismatch previously described. A study in mice has shown that offspring of dams fed a relatively modest 70% nutrient restricted diet developed pronounced weight gain and adiposity in response to a 60% animal fat diet when compared to control animals (Yura et al., 2005). This was postulated to be due to premature leptin surge altering energy regulation by the hypothalamus. Protein restriction has also been utilised in animal models to assess the role of under-nutrition *in utero*. One study in rats showed that offspring of dams fed a 50% protein restricted, isocaloric diet showed increased appetite for energy dense food in early life when compared to control animals (Bellinger et al., 2004). Another study in rats showed that offspring of dams fed an 8% protein diet exhibited increased insulin resistance in later life when compared to pups of dams fed a 20% protein diet (Petry et al., 2001). Taken together these findings suggest that under-nutrition *in utero* programmes the development of a metabolic syndrome-like phenotype in animals (Remacle et al., 2011, Rinaudo, 2012).

1.2.5.2 Over-nutrition *in utero*

Under-nutrition *in utero* has been shown to exert an effect in animal models, however with the increasing dietary intake of mothers a more pressing contemporary societal problem is over-nutrition *in utero*. As such, a number of animal studies have assessed the effects of maternal over-nutrition and obesity on later metabolic health in the offspring. One such study in mice found that offspring of dams fed a palatable obesogenic diet (16% fat, 33% sugar) exhibited increased adiposity alongside cardiovascular and metabolic dysfunction when compared to offspring of dams fed a control diet (3% fat, 7% sugar) (Samuelsson et

al., 2008). Similar findings have been reported in rat models of over-nutrition *in utero* (Nivoit et al., 2009). Another similar study in rats showed that offspring of dams fed a 'junk food' diet high in fat, sugar and salt developed increased adiposity as well as increased plasma glucose and raised circulating fats when compared to the offspring of control dams (Bayol et al., 2008).

1.2.5.3 The early postnatal period - suckling

In terms of developmental origins of obesity risk, as in human studies, studies in animal models have suggested that the early postnatal period may represent a critical time window (Oosting et al., 2010). Although there are a large number of studies into the effects of early life nutrition on later risk of disease (Prior et al., 2011, Symonds, 2010), most of these studies do not draw a distinction between the pregnancy and suckling period. This is due to the fact that the dam is usually fed the intervention diet up until weaning of the offspring. Maternal diet during this period may be important, as this stage of early life has been termed a critical window in the programming of later disease. There is also evidence in rats that a maternal high fat diet during the suckling period results in increased blood pressure, as well as hyperinsulinemia and increased adiposity in normal diet/high fat diet cross fostered offspring (Khan et al., 2005). This study therefore also lends support to the predictive adaptive responses theory mentioned previously (Gluckman et al., 2005).

Studies have shown that increased nutrition and growth during the suckling period is associated with increased obesity later in life (Aubert et al., 1980, Faust et al., 1980, Ozanne et al., 2004), whereas reduced nutrition and growth during this time window permanently reduced weight gain (Cripps et al., 2009, Jimenez-Chillaron et al., 2006) and conferred resistance to diet-induced obesity (Ozanne et al., 2004).

1.2.5.4 Molecular mechanisms – epigenetics?

Given the wealth of evidence from both human and animal studies, developmental programming in early life represents a plausible mechanism in the development of obesity. The process through which it exerts its effects is less clear although epigenetic mechanisms have been suggested as a possible candidate (Waterland, 2005).

1.3 Epigenetics as a mechanism underlying developmental programming

1.3.1 Introduction to epigenetics

Epigenetics, literally meaning 'outside of conventional genetics', is a term used to describe the study of heritable changes in gene expression potential that cannot be explained by changes in DNA sequence (Berger et al., 2009). Epigenetic processes play a key role in development and differentiation and can be influenced and modified by environmental exposures (Jirtle and Skinner, 2007). The term epigenetics refers to the modification of either the DNA itself or proteins that interface with the DNA. These epigenetic modifications are 'sensed' by downstream functional proteins, exerting biological effects such as differential gene expression (Jaenisch and Bird, 2003). The most common epigenetic modifications are histone modification and DNA methylation, the specifics of which are described below.

1.3.2 Histone modification

1.3.2.1 Chromatin and histones

In eukaryotic organisms, genomic DNA is packaged by histone and non-histone proteins to form chromatin (Wolffe, 1998). Each unit of chromatin, also known as a nucleosome, contains 146 base pairs of DNA wrapped around an octamer of four core histones (H2A, H2B, H3, and H4) (Campos and Reinberg, 2009). These histones, particularly their N-terminal tails, are subject to a range of post-translational modifications (Bannister and Kouzarides, 2011). These modifications are implicated in the control of gene expression and therefore genome function (Spencer and Davie, 1999). There are a number of different modifications which occur, including histone phosphorylation, methylation and acetylation (Figure 1.4) (Grant, 2001).

1.3.2.2 Histone acetylation

Histone acetylation is the best characterised of the histone modifications (Vaissiere et al., 2008). It refers to the modification of residues that are acetylated by histone acetyltransferases (HAT) (Kuo and Allis, 1998). This modification is not a permanent one however and can be reversed by histone deacetylases (HDAC) (Richon et al., 2000). Consequently these antagonistic enzymes act as activators and repressors of transcription.

Promoter specific histone acetylation is viewed as a key mechanism in the control of the transcription of individual genes (Deckert and Struhl, 2001). Lower levels of histone acetylation have been linked to an additional epigenetic modification, DNA methylation (An, 2007), which can lead to more permanent gene silencing.

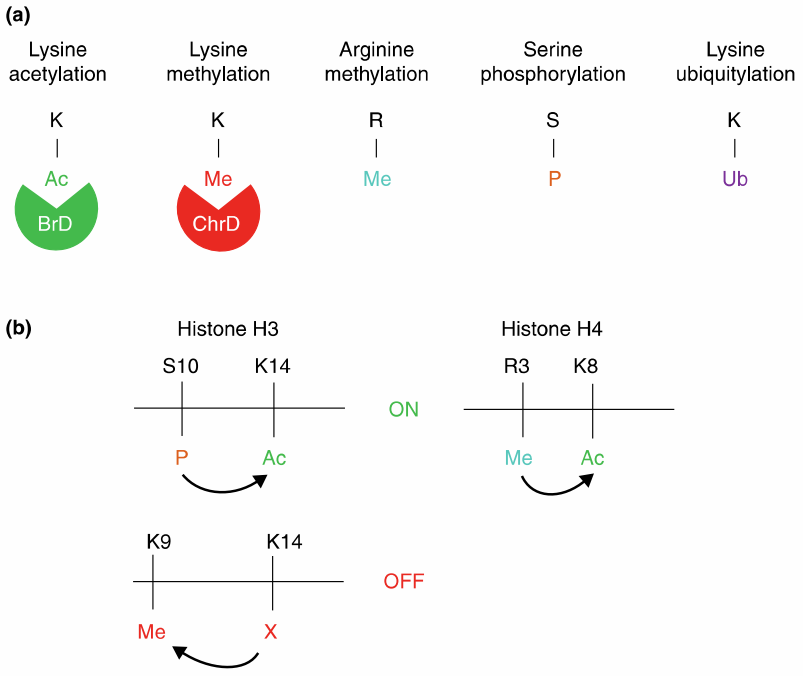


Figure 1.5 Covalent histone modifications.
(a) Types of modifications including acetylation at Lys (K), phosphorylation at Ser (S), methylation at Arg and Lys (R and K) and ubiquitylation at Lys (K). (b) Patterns of modifications. Pairs of modifications, and the sequence of the alterations, correlate with either active or repressed transcription (Grant, 2001).

1.3.2.3 Histone methylation

Histones are also subject to methylation targeted to arginine or lysine residues. Histone arginine methylation has been implicated in gene activation and is involved in the recruitment of methylase enzymes to promoter sequences to act as coactivators (Bauer et al., 2002). These enzymes include coactivator-associated arginine methyltransferase 1 (*CARM1*) that catalyses the transfer of a methyl group to arginine residues in histone H3 (Ma et al., 2001) and protein arginine N-methyltransferase 1 (*PRMT1*) that targets histone H4 (Strahl et al., 2001).

1.3.2.4 Histone phosphorylation

Phosphorylation of Ser-10 on histone H3 has been implicated as an important epigenetic modification in both transcriptional activation and chromosome condensation during mitosis (Wei et al., 1999). This is consistent with conformational changes that occur during these two processes, as chromatin is 'closed' during mitosis and 'open' during transcription

(Gurley et al., 1978). Studies have shown that histone phosphorylation plays a role in the transcriptional induction of key genes in mammalian cells (Cheung et al., 2000). A number of kinases have been implicated in modulation of histone phosphorylation, including the Rsk/Msk families and Snf1 (Hauge and Frodin, 2006).

1.3.2.5 Histone ubiquitination

Ubiquitination has been implicated in a number of key processes within the cell including protein degradation (Lecker et al., 2006), DNA repair (Bergink and Jentsch, 2009) and control of the cell cycle (Nakayama and Nakayama, 2006). Until recently however the implications of ubiquitination in terms of transcription were unknown. Studies have shown that ubiquitin ligase UBR2, one of the recognition E3 components of the N-end rule proteolytic pathway, localises to meiotic chromatin regions, and mediates transcriptional silencing via the ubiquitination of histone H2A (An et al., 2010).

1.3.3 DNA methylation

1.3.3.1 5′-methylcytosine

Methylation of cytosine residues is the most abundant endogenous modification of DNA in mammals. It occurs through the enzymatic addition of a methyl group to the carbon-5 of cytosine (Figure 1.5) (May, 2010). The majority of 5′-methylcytosine (5-mC) in mammalian DNA is present in the form of 5′-CpG-3′ dinucleotides. Some non-CpG sequences such as 5′-CpNpG-3′, 5′-CpA-3′ and 5′-CpT-3′ may also exhibit methylation, however these modifications occur at a much lower frequency (Clark et al., 1994). Recently there has also been an upsurge in interest in 5' hydroxy-methylcytosine, the addition of a hydroxy group to 5-mC, which has been implicated in the control of cell differentiation in embryonic stem cells (Wu et al., 2011a).

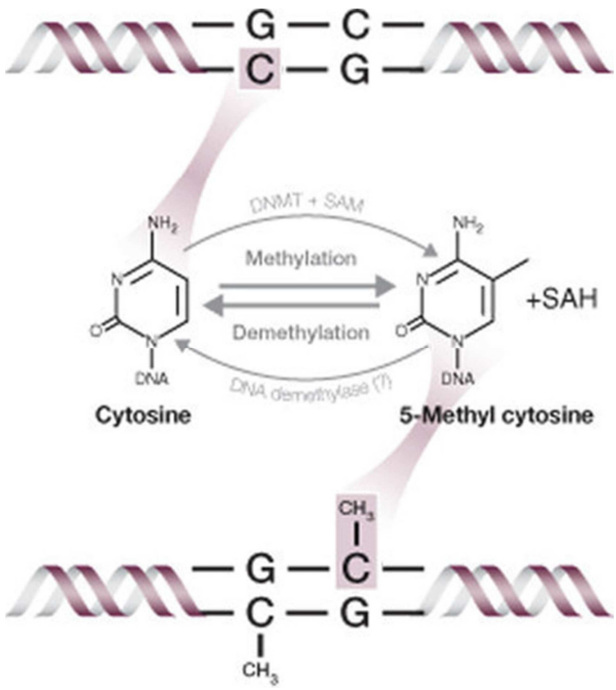


Figure 1.6 DNA methylation.
The 5-carbon cytosine is modified by DNMTs, catalysing the conversion of S-Adenosyl methionine (SAM) to S-Adenosyl-L-homocysteine (SAH). This reaction is possibly reversible. (May, 2010).

1.3.3.2 CpG Islands

CpG dinucleotides are not distributed equally throughout the genome. In 98% of the genome, there is approximately one CpG site per 80 bases. However in the remaining 2% of the genome known as CpG islands (CGIs), CpG density is about five times the level found in the rest of the genome (Deaton and Bird, 2011). CGIs range from about 200 base pairs (bp) to several thousands in length and it has been estimated that there are around 29,000 in the genome (Bernstein et al., 2007). In the majority of cases CGIs are found in gene promoter (Ioshikhes and Zhang, 2000) and exon sequences (Branciamore et al., 2010). These CGIs tend to be unmethylated whereas most CpG sites outside of CGIs tend to be methylated. Some studies suggest that this methylation pattern effectively separates the genome into areas that are to be expressed and areas that are to be transcriptionally repressed (Deaton and Bird, 2011).

1.3.3.3 The interplay of epigenetic modifications

There is a complex interplay between these various modifications, including both DNA methylation and histone modifications that plays a key role in the epigenetic regulation of gene transcription. Some studies have characterised histone modifications as a more transient epigenetic mark than DNA methylation (Barth and Imhof, 2010). Despite this,

they have been implicated in developmental programming in animal models (Lillycrop et al., 2007), suggesting that although the epigenetic mark may be temporary, they may exert a persistent biological effect. However, for the most part, histone modifications have been much less intensively studied in humans. This is largely due to the requirement for prospectively collected chromatin samples and technically demanding laboratory methods. As such DNA methylation is the most commonly studied epigenetic modification in population-based studies. It is this epigenetic mark that this study will focus on.

1.3.4 DNA methylation and development

DNA methylation patterns are established during the development of an organism (Hirasawa et al., 2008). Oocytes generally exhibit lower levels of DNA methylation than sperm cells (Lucifero et al., 2002), however post-fertilisation methylation patterns are erased by a period of genome-wide demethylation during the eight cell stage of blastocyst formation. As embryonic implantation occurs, DNA methylation patterns are re-established (Santos et al., 2002).

1.3.4.1 Maintenance of methylation throughout the lifecourse

In mammals, methylation of cytosine residues is undertaken by three DNA methyltransferase enzymes: *DNMT1* (Robert et al., 2002), *DNMT3A*, and *DNMT3B* (Okano et al., 1999). These enzymes are key functionally, as knockout mice show removal of any one of the genes encoding them is embryonically or postnatally lethal (Li et al., 1992). The DNA methylation pattern is set during early embryonic development, implicating this period as a critical window in the regulation of methylation and therefore gene expression and developmental programming (Dolinoy et al., 2007). The three methyltransferases exhibit differential activity, with DNMT1 preferentially acting on partially methylated DNA. Consequently DNMT1 is considered a maintenance methylase (Robert et al., 2002). In adults DNA methylation levels are highly tissue specific and there is evidence for aging related methylation changes of CpG islands in the promoter of genes (Issa, 2012).

Since a number of developmental processes also involve the erasure of DNA methylation, some studies have suggested that an enzyme with demethylating activity, DNA demethylase protein activation-induced cytidine deaminase (AICDA) can erase DNA methylation (Kangaspeska et al., 2008, De Carvalho et al., 2010). Opponents of this theory

suggest that DNA replication in the absence of methylation maintenance would equally result in passive demethylation (Santos et al., 2002).

1.3.5 DNA methylation as a candidate mechanism

Changing DNA methylation patterns in key elements of the gene, such as promoters and enhancers, can have a profound effect on gene function. Generally speaking, increased levels of methylation result in transcriptional repression (Jones et al., 1998). The best characterised examples of transcriptional repression by DNA methylation in non-disease state are imprinting (Bell and Felsenfeld, 2000) and X chromosome inactivation (Csankovszki et al., 2001). A number of mechanisms have been proposed through which DNA methylation may exert an effect on gene transcription (Figure 1.6) (Bird, 2002). The first theory suggests that the presence of methylated CpG sites within key recognition sequences inhibits transcription factor binding (Kass et al., 1997). A second theory involves the protein complexes Methyl-CpG binding proteins 1 and 2 (MeCP1 and 2) (Boyes and Bird, 1991). These complexes exhibit specificity in binding to methylated CpG sites and as a result can limit access to the regulatory element, inhibiting transcription factor binding. This transcriptional inhibition effect is controlled by the capacity of these MeCPs to recruit the histone deacetylase enzymes (HDACs) (Bird and Wolffe, 1999). HDACs allow histones to interact by deacetylating lysine residues in their N-terminal tails. This results in the formation of conformationally 'closed' chromatin, which acts as a repressor of transcription. A range of other factors have been implicated in the mediation of this process including methyl binding domains (MBD) however their regulatory role has yet to be fully elucidated (Fraga et al., 2003).

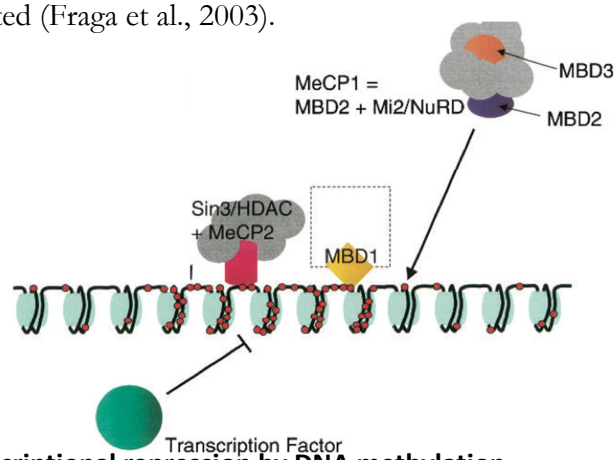


Figure 1.7 Transcriptional repression by DNA methylation.
DNA strand shown with methylated CpGs in red. Below (green) is a transcription factor unable to bind to its recognition site due to the presence of methylated CpGs. Above are protein complexes that bind methylated CpGs, including methyl-CpG binding protein (MeCP2) and Sin3-HDAC, a complex of the corepressor Sin3 and histone deacetylase. Also present is MeCP1, comprising of MBD2 plus NuRD corepressor complex. MBD1 (yellow) has yet to be fully characterised (Bird, 2002).

1.3.6 Evidence for epigenetic mechanisms mediating the influence of early life exposures on the risk of obesity

Changes in epigenetic patterning, and particularly changes in the level of DNA methylation can have a profound effect on an individual's phenotype (Hitchins et al., 2007), although this evidence is not widespread outside of imprinting disorders and cancer at the current time. As outlined earlier, there is an increasing body of evidence linking dietary and environmental exposure *in utero* to epigenetic changes in offspring: These interventions include nutritional insult *in utero* (Chmurzynska, 2010), depletion of maternal folate levels *in utero* (McKay et al., 2004), as well as maternal environmental exposures such as polycyclic aromatic hydrocarbons (PAHs) as found in cigarette smoke (Perera et al., 2009). However, regarding these various exposures, there is currently very little published literature directly linking changes in exposure with altered epigenetic patterning and in turn to the development of an obesity or metabolic disease phenotype.

Epigenetic modifications provide a very attractive mechanism whereby early life exposures are 'captured' by the genome and exert effects on gene expression and health in later life. Animal models have characterised differential expression of genes in offspring exposed to over-nutrition *in utero* and which subsequently go on to develop an obese phenotype (Jimenez-Chillaron et al., 2009). These differential levels of expression may be regulated by epigenetic processes; indeed many of the appetite regulatory genes are rich in CGIs or subject to histone modifications, suggesting that this may be the case (Widiker et al., 2010, Stevens et al., 2009). CGIs in the promoter regions of these key appetite regulatory genes may be subject to differential levels of methylation. Increased levels of DNA methylation around the transcription start site could conceivably alter levels of gene transcription and therefore alter appetite modulation. With obesity as a known pathology of energy homeostasis it does not require much extrapolation to make the link between chronically increased appetite levels, increased energy intake and the development of increased adiposity. It is clear therefore that the levels of DNA methylation within CGIs located in genes key to appetite regulation may be able to inform us about an individual's risk of developing obesity in later life. Many other pathways leading to obesity, including impaired glucose and lipid metabolism and adipogenesis, could also be plausibly perturbed by epigenetic alterations. This hypothesis forms the basis of this thesis.

1.4 Studying epigenetic mechanisms in the early life programming of obesity

1.4.1 Animal studies

The wealth of literature using animal models to investigate the field of developmental programming (Langley-Evans, 2001, Begum et al., 2012, Sinclair et al., 2007), shows us that they provide certain advantages over human models (Vickers et al., 2005). One advantage is that nutritional interventions and other more stringent insults that are not plausible in humans can be relatively easily implemented in animals (Lo et al., 2011). More practically, the use of laboratory animals whose environmental exposures are carefully controlled can allow an environmental insult to be causally linked to a phenotypic change. The shorter life span of rodents in particular also offers an advantage in animal models, in that it is possible to implement time course experiments and follow animals throughout their development, taking phenotypic and genotypic measurements where required (Symonds, 2010). Animal models also allow a large range of tissue types to be assessed for molecular changes as the result of an exposure, something which is very difficult to achieve in a human study (Dzamko et al., 2010). However an inherent problem with animal studies is the fact that, although acting as a useful surrogate, animal models cannot act to replace human investigations. There will always be questions about the transferability of any molecular or phenotypic change recorded in an animal model to a human equivalent (Garland et al., 2011). Consequently, although proving useful as a tool to investigate interventions which are implausible in a human study, animal models should be used in conjunction with human studies in order to allow interesting findings to be validated.

1.4.2 Bioinformatic approaches

Bioinformatics is the application of computer science to the information technology to the field of biology and medicine. It allows us to assess the genome in order to identify more suitable candidate genes for methylation analysis, as well as informing the most relevant places within genes to assess for differential DNA methylation. There are a number of methodological approaches which can be used to identify potentially differentially methylated loci, including a range of data mining techniques. One methodology which can benefit from in-depth bioinformatic analysis is the gene expression microarray. When differential gene expression is detected as a result of a common complex disease

phenotype, such as obesity, it is possible to use bioinformatic analyses to identify genes within the differentially expressed subset which are likely to undergo differential DNA methylation. Bioinformatic approaches for the prioritisation of epigenetic target genes have previously been described in the context of differential gene expression response following nutritional exposure *in utero* in mice (McKay et al., 2008), and have been assessed in relation to BMI in children (Turcot et al., 2012). Both of these studies integrated a number of different bioinformatic tools, producing a workflow with which to identify target genes for downstream analysis. The first step in this kind of workflow is to set a p-value FDR (false discovery rate) cut off between the two experimental groups (Chumbley and Friston, 2009). This is commonly set at 0.05. Following this step the remaining genes in the workflow can be assessed to discern if they are linked to the pathology of interest. This can be performed using a range of tools including Genomatix LitInspector which text data mines literature for key word of interest (Frisch et al., 2009). It is also possible to map the genes that have progressed this far through the workflow to ontology pathways of interest using a pathway analysis tool such as DAVID or Ingenuity Pathway Analysis (Berisha et al., 2011). Once a subset of genes has been identified which map to the pathology of interest, it is then possible to assess how likely the remaining loci are to be differentially methylated. This step can use a number of tools including MeInfoText (Fang et al., 2011). The next stage of any such analysis would be to select which region of the gene is of interest. Promoter and inducer sequences can be mapped using a range of online tools including Genomatix (Palou et al., 2011). The sequence of interest, once identified, can then be assessed for the presence of CpG islands using a number of tools including CpG Island Explorer (Ongenaert and Van Criekinge, 2005). At this stage it is possible to assess the localisation of any SNPs within the sequence using tools such as the NCBI sequence viewer. This then allows the gene of interest to be fed into a downstream platform for DNA methylation analysis. A summary of a range of commonly used bioinformatics tools is presented in Table 1.1 (Laird, 2010). Bioinformatic analysis of expression datasets is dealt with in more detail in Chapter 3.

Resource	Purpose
Batman	MeDIP DNA methylation analysis tool
BSMAP	Whole-genome bisulphite sequence mapping
CpG Analyser	Windows-based program for bisulphite DNA
CpGcluster	CpG island identification
CpG Island Explorer	Online program for CpG Island identification
CpG PatternFinder	Windows-based program for bisulphite DNA
CpG Promoter	Large-scale promoter mapping using CpG islands
CpG ratio and GC content Plotter	Online program for plotting the observed:expected ratio of CpG
CpGviewer	Bisulphite DNA sequencing viewer
CyMATE	Bisulphite-based analysis of plant genomic DNA
EMBOSS CpGPlot	Online program for plotting CpG-rich regions
Epigenomics Roadmap	NIH Epigenomics Roadmap Initiative homepage
Epinexus	DNA methylation analysis tools
Gentomatix	Software suite including promoter searching
Ingenuity	Pathway analysis and nomenclature software
MEDME	Software package (using R) for modelling MeDIP experimental data
methBLAST	Similarity search program for bisulphite-modified DNA
MethDB	Database for DNA methylation data
MeInfoText	Text data mining for methylated genes
MethPrimer	Primer design for bisulphite PCR
methPrimerDB	PCR primers for DNA methylation analysis
MethTools	Bisulphite sequence data analysis tool
MethyCancer Database	Database of cancer DNA methylation data
Methyl Primer Express	Primer design for bisulphite PCR
Methylumi	Bioconductor package for analysing DNA methylation data from Illumina platforms
Methylyzer	Bisulphite DNA sequence visualization tool
mPod	Genome-wide DNA methylation viewer integrated with the Ensembl genome browser
PubMeth	Database of DNA methylation literature
QUMA	Quantification tool for methylation analysis
TCGA Data Portal	Database of TCGA DNA methylation data

Table 1.1 Bioinformatic resources

BSMAP, Bisulphite Sequence Mapping Program; CyMATE, Cytosine Methylation Analysis Tool for Everyone; EMBOSS, European Molecular Biology Open Software Suite; MeDIP, methylated DNA immunoprecipitation; MEDME, Modelling Experimental Data with MeDIP Enrichment; NIH, US National Institutes of Health; QUMA, Quantification Tool For Methylation Analysis; TCGA, The Cancer Genome Atlas (Laird, 2010).

1.4.3 Human studies - Epidemiological approaches

1.4.3.1 Introduction

It is possible to investigate the effect of epigenetic modifications on human disease risk through the use of traditional epidemiological study design methods. These types of studies are useful because they allow us to study the frequency of differential DNA methylation across the population in relation to the development of common complex disease. For example recent studies in a cohort of preterm (n = 121) and term born (n = 6,990) babies showed that methylation of the *TACSTD2* gene was associated with fat mass in slow versus rapid growing infants (Groom et al., 2012). Another example of a recent epidemiological study linking DNA methylation to an obesity related outcome was performed in a cohort of 25 overweight or obese men who were participating in an eight week energy restriction intervention. DNA methylation levels in several genes were significantly modified as a result of the hypocaloric diet (Milagro et al., 2011). However there are a number of issues which need to be considered when designing this type of study when considering their application to the investigation of epigenetic programming of obesity (Heijmans and Mill, 2012).

1.4.3.2 Tissue specificity

DNA methylation patterns are highly tissue specific (Byun et al., 2009). This issue is further complicated by the fact that the great majority of tissue types of interest can only be collected in living human cohorts with invasive procedures such as biopsies. This issue is solved in animal models where the tissue of interest can be retrieved and processed *post mortem* (Zeng et al., 2011). In human studies, however, it is more conventional to select a source of DNA that is readily available and non-invasive in the form of either blood or saliva (Tierling et al., 2011). In this instance epigenetic patterns can only provide a surrogate for the target disease tissue of interest.

1.4.3.3 Temporality of epigenetic change

In an individual with a disease phenotype, any changes in DNA methylation observed may be the cause of the disease rather than an antecedent. In an ideal study design tissue samples for DNA methylation analysis would be taken prior to the onset of disease to rule out reverse causation. However in most instances studies tend to focus on diseased individuals. Prospective sampling is more likely in long-term studies that take samples of

readily available tissues such as blood and saliva at multiple time points as study members are followed throughout their lifecourse.

1.4.3.4 Study design

A key step in the implementation of an epidemiological study is the choice of study design. The choice of study design depends on many factors including the exposure and outcome of interest and adjustment for confounding factors. In this section the various study design options are considered.

Cross-sectional study

In a cross-sectional study, all of the factors that are of interest in the study population are measured at the same time. This study design is best suited to instances when the biological measurement is discrete and the population is well defined (Kestenbaum, 2009). A study design of this type investigating DNA methylation could, for example, measure gene specific methylation at a number of loci in a cohort of aged individuals with specific phenotypic traits (see Chapter 6). A cross-sectional design also allows DNA methylation levels between two subsets of the population to be compared, for example the results could be stratified for sex or age (Fenech et al., 1998). This type of study design is useful for smaller scale studies where hypothesis generation is the preferred outcome, however to infer causality between differences in DNA methylation levels and a disease outcome other approaches are required.

Retrospective case-control study

In a case-control study, individuals displaying a disease phenotype (cases) are matched to individuals free of disease (controls) from the same population (Schlesselman and Stolley, 1982). A comparison of the two groups with respect to specific exposure information can identify the relative risk of disease. Control selection is a key step in this study design as bias and confounding can result in inappropriately matched controls (Geneletti et al., 2009). A case-control study design into DNA methylation would involve the assessment of gene-specific or genome-wide methylation in samples taken from groups of cases and controls (Moore et al., 2008). Again, like the cross-sectional study, it is not possible to infer causality from a case-control study regardless of the biological measurement taken.

Cohort study

In a cohort study, individuals are recruited to participate initially over a defined period, be it

weeks, months or years (Breslow and Day, 1987), however in practise many cohort studies are open-ended and will follow participants throughout their life. When the individuals first attend clinic, baseline samples are taken to allow any biomarkers of interest to be monitored. Study members are then followed up at a number of time points after the initial study visit, when additional biological and clinical measurements are collected (Cao et al., 2009). This allows any change in phenotype or, for example, DNA methylation level to be assessed over time. Cohort studies are typically large, with hundreds to thousands of participants. If followed over time this type of study referred to as a longitudinal cohort study. The strength of this type of study is that clinical and biological measurements have been collected in all individuals prior to the development of the disease phenotype. This allows any biological differences present in the diseased individuals to be assessed in relation to their earlier measurements and thus strengthens causal inference. This study design is particularly useful in studies of DNA methylation as the levels prior to the development of disease can be assessed, removing the problem of reverse causation found with other study types (Brooks et al., 2010).

Birth cohort

Birth cohorts enrol individuals *in utero* or in very early life by recruiting their parents during pregnancy or in the early postnatal period (Wadsworth, 2002). These cohorts are particularly useful when assessing early life influences on later disease risk, as exposures during the 'critical windows' of *in utero* and early postnatal life can be measured (Ponsonby et al., 2011). Biological samples are often available in these individuals and in some cohorts tissues available from both mother and child in the form of cord blood, placenta and saliva. This allows DNA methylation status as a very early age to be discerned (Terry et al., 2008). However, many birth and longitudinal cohort studies have collected biological samples at one time point only, often in adulthood, despite having extensive measures on the cohort from an early age. As individuals, ideally both parent and child, can be followed up over time, disease phenotypes and epigenetic changes can be monitored longitudinally over this period (Flom et al., 2011).

Nested case-control study

Nested case-control studies use a subset of individuals selected from a cohort study. Individuals who develop the disease outcome of interest during the cohort study period are selected for further analysis, alongside matched disease-free control individuals (Sedgwick, 2010).

Intervention study/Randomised controlled trial

Intervention studies, including randomised controlled trials, allow the effect of a modification in behaviour or environment on an individual's health to be investigated. Such interventions in human studies can include either a lifestyle-based modification, such as increased exercise (Rosenstock et al., 2010), or a dietary intervention. Dietary intervention studies have been conducted to investigate the influence of specific factors on DNA methylation. For example, studies have shown that supplementation with folate can exert an effect on DNA methylation levels (Pilsner et al., 2009a).

Family-based study

This type of study recruits multiple family members, often structured around parent and offspring triads (Laird and Lange, 2006). This allows the heritability of any biological changes in the child to be mapped to those in both parents. The heritability of epigenetic changes including DNA methylation has yet to be fully elucidated but the persistence of a small proportion of DNA methylation marks across generations is plausible and may contribute to the intergenerational transmission of obesity risk (Whitelaw and Whitelaw, 2008).

1.4.3.5 Use of appropriate study design

The choice of study design is dependent upon the question being addressed, for example whether prediction, prevention or treatment is the key target of epigenetic investigation. In the context of developmental programming longitudinal cohort studies, preferably with parent and offspring sampling and data available would be the ideal. The relative merits of different study designs are summarised in Table 1.2.

Study design	Possible findings	Strengths	Limitations
Cross-sectional	Prevalence of DNA methylation in a well- defined subgroup	Easy to implement and analyse	Small numbers and no causality
Retrospective case-control	Comparing DNA methylation between individuals with and without disease	Well defined and large phenotypic change	No causality unless prospective
Cohort	DNA methylation as a disease risk factor	Maps trends across population	Can be unwieldy – large numbers
Nested case-control	DNA methylation as a biomarker of early disease	Allows early disease to be mapped and tracked	Smaller numbers so less well powered
Intervention	Effect of intervention on DNA methylation levels	Clear point of biological inducement	Limited interventions
Family-based	Transgenerational inheritance of DNA methylation	Data on parental phenotype and epigenotype	Analysis of sex specific effects
Birth cohort	Influence of *in utero* and early postnatal factors on establishment of DNA methylation levels	Early life measures of exposure	Costly to follow up for long period

Table 1.2 Study designs and relative merits

1.4.4 Epigenetic methods to quantify DNA methylation

The methods available for the appraisal of variation in DNA methylation are numerous. Discussion here is limited to those approaches adopted in this thesis. A broader appraisal of the advantages and disadvantages of the more commonly applied methodologies is provided in Table 1.3 (Laird, 2010).

Technology	Features							Potential sources of bias							
	Unambiguous identification of CpG measured	In *cis* co-methylation information	Non-CpG methylation information	Allele-specific measurement capability	Good coverage of regions with low CpG density	Compatible with low amounts of input DNA	Full repeat-masked genome coverage	Copy-number variation bias	Fragment size bias	Incomplete bisulphite conversion bias	Bisulphite PCR bias	Cross-hybridization bias	DNA methylation status bias	GC content bias	CpG density bias
Infinium	(•)					•				•	•	•			
Enzyme–chip	(•)	(•)			(•)				•			•		•	
MeDIP–chip							•	•				•		•	•
BSPP	•	•	•	•						•	•		•		
BC–seq	•	•	•	•						•	•		•		
RRBS	•	•	•	•		•				•	•				
Enzyme–seq	•	•		•	(•)	•			•						
MeDIP–seq				•			•	•						•	•
WGSBS	•	•	•	•	•	•	•			•	•				

Table 1.3 Features and sources of bias for DNA methylation analysis techniques
'•' indicates that the method has this feature or potentially has this bias; '(•)' indicates that the method has this feature to a limited extent or in some circumstances. BC-seq, bisulphite conversion followed by capture and sequencing; BSPP, bisulphite padlock probes; –chip, followed by microarray; MeDIP, methylated DNA immunoprecipitation; RRBS, reduced representation bisulphite sequencing; –seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing (Laird, 2010).

1.4.4.1 Gene specific analyses

Bisulfite sequencing

Bisulphite sequencing has for many years been the gold standard for assessing methylation status of DNA. In common with many methods, the DNA is treated with sodium bisulphite that acts to deaminate un-methylated cytosine residues in the sequence into uracil (Grunau et al., 2001). Methylated cytosine residues however remain resistant to this modification and persist. The sequence is then amplified and the uracil bases amplified as thymine residues. Subsequent sequencing of the DNA samples provides information on the methylation status of each CpG site by comparing the relative abundance of cytosine and thymine residues (Deng et al., 2009). Bisulphite sequencing has been widely applied, its main advantage being that it identifies the methylation status at base pair resolution within the target sequence; it can however be a labour intensive and costly process. The principal disadvantage of this method is that the methylation levels cannot be quantified (although this can be overcome by highly parallel next generation sequencing where high fold coverage is obtained) limiting the ability to draw associations with any transcriptional consequences (Zilberman et al., 2006). Bisulfite sequencing can be applied in target specific manner – that is small fragments of the genome can be analysed rather than profiling the genome in its entirety.

Pyrosequencing

One such targeted approach is pyrosequencing. This method is based on the PCR amplification of bisulphite modified DNA with primers specific to an amplicon of interest and adopts a sequence by synthesis approach. One of the primers is biotin labelled, which allows the PCR product to bind to streptavidin coated sepharose beads (Colella et al., 2003). A series of wash steps denatures the amplicon and washes the non-biotinylated strand off the beads. The remaining bound strand is then transferred to a specialised pyrosequencing reaction plate before reaction buffers, a sequencing primer and DNA synthase enzyme are added. One nucleotide at a time is then added to the reaction following a predetermined dispensation order (Vasiljevic et al., 2011). If the dispensed base is complementary to the DNA strand this nucleotide is incorporated into the sequence by DNA synthase, releasing pyrophosphates that are converted via an enzyme cascade into light. This light is then quantified by a sensitive camera. The light emitted is proportional to the amount of nucleotide incorporated into the sequence and as such the ratio of light emitted following the dispensation of Cytosine or Thymine at the CpG site represents the level of methylation at the site (Tost and Gut, 2007). The advantages of this protocol are that it is quantitative, reproducible, and custom assays are relatively straightforward to develop. This method has been applied widely in the work presented in my thesis and further details are provided in the relevant Results chapters.

1.4.4.2 Global analyses

Repetitive element assays

This type of assay focuses on repetitive elements of the genome; transposable sections of DNA that share the same sequence and have a variety of functions. These repeat elements include long interspersed nuclear elements (LINEs) (Fryer et al., 2009), short interspersed nuclear elements (SINEs) (Arnaud et al., 2000) and *Alu* and *Sat2* repeats (Gao et al., 2011b). assays in this category generally utilise PCR primers specific to one of these repeating elements following bisulphite modification (Yang et al., 2004). This allows DNA methylation within the repetitive element to be assess by one of the sequencing methodologies previously discussed (Weisenberger et al., 2005). The methylation measure that is derived from this analysis is a surrogate for global methylation levels as it is an average of methylation in the repetitive element found throughout the genome (Baccarelli et al., 2010a). It must be remembered however that methylation within repetitive elements may have functional relevance to the specific element of interest but may not necessarily be

related to gene expression. LINE-1 global DNA methylation analysis was conducted as part of the work presented in my thesis.

Immunohistochemical approaches

Immunohistochemistry can be used to assess DNA methylation at a tissue and even cell specific level. Using primary antibodies specific to 5-Methylcytosine or related proteins such as the Methyl binding domain molecules (MBDs), it is possible to stain sections of tissue for the presence or absence of DNA methylation (Yang et al., 2011). When combined with fluorescence-conjugated secondary antibodies it is therefore possible to quantify the levels of DNA methylation within tissue sections (Brown et al., 2008). Further details of this approach to are provided in Chapter 3.

1.4.4.3 Genome-wide analyses

Increasingly, more and more epigenetic studies are performed at a genome-wide scale. Platforms available include microarray based approaches and sequence based approaches, both of which are undergoing rapid development and refinement. Neither has been applied in the thesis work presented here.

1.4.5 Epigenetic marks as biomarkers of disease susceptibility

1.4.5.1 Evidence in cancer

DNA methylation represents a potential early biomarker of disease risk. There is a wealth of evidence in cancer to demonstrate profound shifts in DNA methylation patterns in tumour tissue (Hinoue et al., 2012, Gu et al., 2010) and these biomarkers may have utility in a diagnostic or prognostic context. One recent study in colorectal cancer showed that increased levels of methylation in the key regulator genes *FBN2* and *TCERG1L* were associated with decreased gene expression. Both of these genes also showed a high level of DNA methylation in a range of colon cancer types including adenoma and carcinomas but not in normal colon tissue (Yi et al., 2012). A recent case-control study in prostate cancer patients has also show a range of DNA methylation markers in urine were able to discriminate between diseased and non-diseased individuals. (Payne et al., 2009). There has also been evidence in lung cancer that increased DNA methylation levels in the *SHOX2* gene in bronchial fluid can be used as a diagnostic test for lung cancer risk (Dietrich et al., 2011). There are however, fewer examples of the application of DNA methylation patterns in non-tumour tissue as an early biomarker of disease (Hsiung et al., 2007). Peripheral

blood cell DNA has also been found to be a sensitive epigenetic biomarker of disease risk in ovarian and bladder cancers (Marsit et al., Teschendorff et al., 2009), although these results require replication (Relton and Smith, 2012).

1.4.5.2 Evidence in other chronic diseases

Aside from these findings in cancer there have also been recent data on DNA methylation as a biomarker of a number of other conditions. A recent study has shown that DNA methylation levels across a total of 349 CpG sites was associated with the presence and severity of chronic obstructive pulmonary disease (COPD) (Qiu et al., 2011). There has also been recent evidence suggesting that DNA methylation levels are a potential biomarker for emphysema (DeMeo et al., 2011). Taken together these results suggest that DNA methylation may represent a useful diagnostic marker in non-neoplastic lung disease. There is however increasing evidence in that DNA methylation could play a similar role in the prediction of obesity and related sequelae like type 2 diabetes. Two recent studies have reported an association of DNA methylation at birth and later childhood adiposity (Relton et al., 2012, Godfrey et al., 2011b).

1.4.5.3 Evidence in obesity-related sequelae

The role of DNA methylation as a biomarker in obesity related sequelae is not as well defined as that in cancer. There is however increasing evidence in a number of conditions suggesting that DNA methylation may play a role. In addition to the evidence alluded to above linking DNA methylation patterns in early life with later childhood body composition (Godfrey et al., 2011b, Relton et al., 2012) There is also an increasing body of evidence suggesting that DNA methylation signatures may represent a potential biomarker of obesity risk. One such study showed that low energy diet-induced weight loss in obese humans altered the DNA methylation status of the ATP10A and CD44 genes. This suggests that DNA methylation measurements taken at baseline may prove useful as predictive markers of weight loss (Milagro et al., 2011). Another study has shown that, in blood leukocytes, DNA methylation levels in the genes UBASH3A and TRIM3 were decreased in obese individuals when compared to lean controls (Wang et al., 2010a). Additionally, a recent study in a Singapore cohort reported that increased global DNA methylation measured by *Alu* and *Sat2* repetitive element assays was positively associated with increased prevalence of cardiovascular disease (Kim et al., 2010), whilst another study has shown that hypomethylation at loci susceptible to prenatal environmental insults was associated with risk of myocardial infarction (Talens et al., 2011). Another recent study

showed that insulin resistance measured by HOMA was strongly associated with increased global DNA methylation measured by the *Alu* assay (Zhao et al., 2011). Studies in our laboratory have also identified differential methylation patterns in the type 2 susceptibility gene *KCNQ1* associated with insulin sensitivity (Elliott et al., 2012).

1.4.5.4 Opportunities in obesity

In summary, there have been some promising findings relating DNA methylation at both a gene specific and genome wide level with respect to the risk of developing obesity and related sequelae. However further work is needed to elucidate whether the observed influences of early life exposures on later obesity are indeed mediated by epigenetic processes. This issue is the focus of my thesis.

1.5 Hypotheses, aims and objectives

1.5.1 Hypotheses

Obesity is a major public health burden and early life exposures may play a role in the development of this condition. DNA methylation is subject to change during critical windows such as the perinatal period, and thus aberrant DNA methylation in response to early life exposures may predispose an individual to the development of obesity. My thesis aimed to address the following hypotheses:

1. The influence of early life exposures upon later obesity (and related sequelae) are mediated in part by epigenetic mechanisms.

2. Differences in epigenetic signatures (namely DNA methylation) associated with early life exposures can be identified in animal studies and human populations.

3. Epigenetic patterns (namely DNA methylation) are associated with obesity related traits in childhood and adulthood.

1.5.2 Aims and objectives

The specific objectives of this project were:

- To identify tissue specific differential DNA methylation in rat hypothalamus in response to nutritional insult.
- To create a bioinformatic workflow allowing differentially methylated target genes to be identified from a gene expression array dataset.
- To quantify differential methylation in target genes identified by the bioinformatic workflow in tissues from an animal model of developmental programming.
- To assess gene specific DNA methylation in relation to markers of metabolic health at age eight years in a human cohort study.
- To assess the effect of global DNA methylation levels at age 50 years on markers of metabolic health in a human cohort study.
- To assess the effect of early life exposures on DNA methylation in later life in human cohort studies.

Chapter 2: An Immunohistochemical Approach to Identify Differential Methylation in the Appetite Regulatory Centre of the Brain

2.1 Background

Obesity results from an imbalance between energy intake and energy expenditure which is sustained over substantial periods of time (Zhang et al., 2008). The brain and in particular the hypothalamus play a key role in the regulation of energy homeostasis (Williams et al., 2001). The arcuate nucleus (ARC) is the section of the hypothalamus which integrates a series of complex and diverse hormonal and nutritional signals (Schwartz et al., 2000) in order to mediate food intake and energy expenditure in physical activity (Coppari et al., 2005, Sainsbury and Zhang, 2010), and therefore energy balance. Hormones such as leptin (Cowley et al., 2001), insulin (Niswender et al., 2003), adiponectin (Guillod-Maximin et al., 2009, Qi et al., 2004), and ghrelin (Hewson and Dickson, 2000) and nutritional signals such as circulating glucose (Wang et al., 2004) and fatty acids (Lam et al., 2005) all exert effects on the ARC. These signals are then integrated by neuropeptides including weight increase inducing peptides such as neuropeptide Y (NPY) (Chen et al., 2004) and agouti-related peptide (AgRP) (Wilson et al., 1999), and weight loss inducing peptides such as proopiomelanocortin (POMC) (Boston, 2001) and cocaine and amphetamine regulated transcript (CART) (Rogge et al., 2008, Wierup et al., 2005). (See Fig. 2.1 and Introduction).

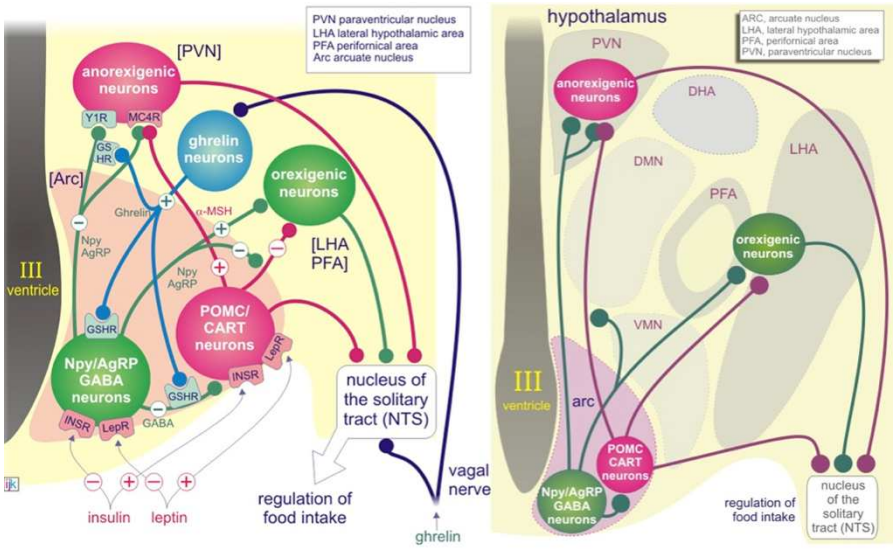


Figure 2.1 Hypothalamic regions in the regulation of appetite.

The role of DNA methylation in this brain region specifically is therefore of interest as an effect on the expression of any of a large number of key genes may result in the perturbation of the mechanisms controlling food and therefore energy intake and also energy expenditure through physical activity. A sustained positive energy balance is likely to result in the development of an obese phenotype.

2.2 Aim and objectives

The aim of this chapter was to assess tissue specific DNA methylation in rat hypothalamus in response to nutritional insult. Given the role of hypothalamic nuclei in the control of appetite, dysregulation of gene expression in these cells may mediate altered feeding habits e.g. increased food intake and/or reduced physical activity and therefore the positive energy balance which is a prerequisite for the development of an obese phenotype. Since DNA methylation is known to regulate expression of a number of genes, the pattern of DNA methylation was explored in hypothalamic tissue in a rat model of developmentally programmed appetite dysregulation in search of evidence of gross shifts in DNA methylation signatures. Immunohistochemical techniques were optimised and applied to the labelling of rodent brain tissue.

2.3 Methods

2.3.1 Animal model

Whole rat brains were kindly provided by Dr. Sue Ozanne at the University of Cambridge. The following is a summary of the experimental protocol for this rodent model of developmentally programmed appetite dysregulation. Adult female Wistar rats were housed individually and were maintained at 22°C on a 12:12-h light-dark cycle. When they reached weight of between 235 and 250 g, they were mated. The day on which vaginal plugs were observed was taken as day 0 of gestation. Dams were fed *ad libitum* either a control diet (containing 20% protein) or an isocaloric low protein (8% protein) diet (both diets were purchased from Arie Blok, Woerden, the Netherlands) during gestation and lactation. Crossfostering techniques were used at birth to establish these study groups: 1) controls [offspring of control dams, culled to eight (four males and four females) and suckled by control dams]; 2) recuperated (offspring of dams fed a low-protein diet during pregnancy,

but nursed by control dams, culled to four to maximize the plane of nutrition); and 3) postnatal low-protein (PLP) (offspring of control dams nursed by low-protein-fed dams, unculled to minimize the plane of nutrition). Plane of nutrition in the context of this study refers to the quantity of food intake per animal. Body weights of animals were recorded at birth and at days 3, 7, 14, and 21 of age. At day 21 pups were removed from dams and starved overnight.

2.3.2 Brain preparation

The whole brain was extracted in a single piece from freshly culled male animals and placed into a rat brain coronal slicer matrix. Two coronal cuts were made in order to isolate the hypothalamus, one at the optic chiasm (9.48mm interaural, 0.48mm Bregma) and one immediately prior to the pons (4.48mm interaural, -4.52mm Bregma) (Figure 2.2). These slices ensure that the portion of the brain isolated contained hypothalamus throughout as coronal sections were taken. The isolated midbrain section was then fixed overnight in 10% neutral buffered formalin before being processed using Newcastle Hospitals NHS Trust Cellular Pathology 'Routine overnight' protocol. This protocol was a 14 hour process with an initial 30 minute exposure to 10% formalin followed by 95% then four 99% xylene steps for one hour under vacuum at 40 degrees Centigrade. A further four hour-long steps with xylene were completed before the tissue was processed to VWR Gurr Fibrowax 36142. Sections were then cut at three microns using a microtome (HM325, Microm, UK) and mounted to Superfrost electrostatically coated slides (VWR, Leicestershire, UK), dried at 60°C for one hour and then overnight at 37°C.

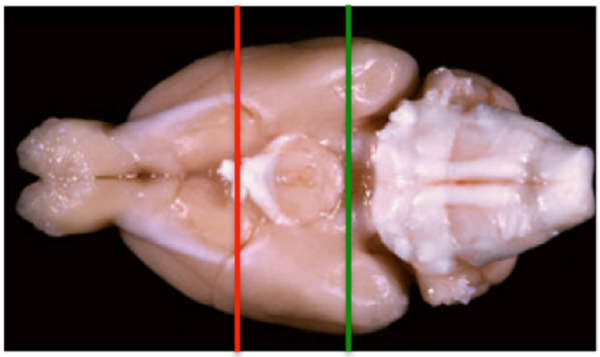


Figure 2.2 Ventral surface of rat brain showing gross cuts made to isolate hypothalamus. The red line indicates the cut made at the optic chiasm (9.48mm interaural) and the green line indicates the cut made prior to the pons (4.48mm interaural).

2.3.3 Immunohistochemistry - Primary antibody selection

Monoclonal [33D3] antibody to 5-Methyl Cytidine (5MC) (Stratech, Suffolk, UK) was selected as a marker for DNA methylation within the nucleus. Monoclonal (22C9) antibody to Neurone specific enolase (NSE) (Novocastra, Newcastle, UK) and monoclonal (2G10) antibody to Neurone specific Beta III Tubulin (B3T) (Abcam, Cambridge, UK) were selected as intracellular neuronal markers.

2.3.4 Immunohistochemistry - Protocol

The complete immunohistochemistry protocol is summarised in Figure 2.3. Slides were dewaxed in Xylene and blocked for 10 minutes with Methanol/peroxide solution. Sections were then washed in water before microwave incubation in Antigen Unmasking Solution (Vector, Peterborough, UK). Runs were attempted with 5, 10 and 15 minute microwave incubations and 10 minutes was selected as the most effective incubation time. Sections were then blocked with 10% Normal Horse Serum (Vector) for 10 minutes before incubation with one of the primary antibodies. Antibodies were diluted in PBS and runs attempted at several concentrations within the range of 1:600 to 1:50 for each antibody. Incubation times were varied from two hours to 30 minutes at room temperature for each antibody and an incubation time of 60 minutes was selected for each of the primary antibodies. 5MC and NSE were both optimised to a dilution of 1:200 in PBS and B3T was optimised to a dilution of 1:500 in PBS. Sections were washed in water before being incubated with ImmPRESS Universal reagent (Vector) for 30 minutes at room temperature. A further wash step was implemented before development for colour using the ImmPact DAB system (Vector) for 10 minutes before counterstaining nuclei with Haematoxylin for one minute. Sections were then dehydrated and mounted under cover slips using DPX mountant. The optimised process is summarised in Figure 2.3. Figures 2.4 and 2.5 show optimisation of staining protocols for 5mC and NSE with NPA (no primary antibody) and gradually increasing concentrations of primary anitbody, from 1:600 to 1:50. In the case of both antibodies the optimal concentration was determined to be 1:200. This was evidenced in the case of 5mC by the clear positive nuclear signal and lack of background in the rest of the slide.

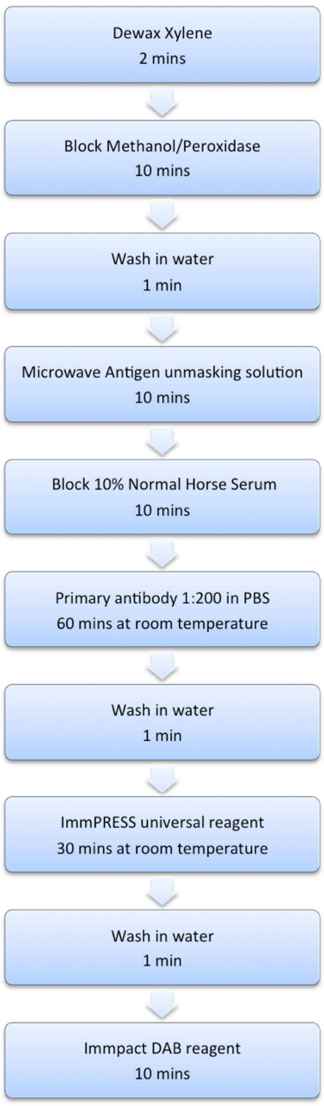


Figure 2.3 Flow diagram of optimised immunohistochemistry process.
Includes incubation times and temperatures where appropriate.

2.3.5 Immunohistochemistry - Selection of control tissues

A number of controls were incorporated into the IHC runs to ensure that the system was working optimally and the antibodies were exhibiting specificity. To ensure optimal antigen retrieval, human colorectal adenomatous polyp biopsies were stained for Ki-67 (MIB-1), a cell proliferative marker that has been used extensively in our lab. This ensured consistency between runs. Since the tissue type for the experimental runs was brain, non-brain was used as a control to ensure that both the 5MC and NSE primary antibodies specifically stained neurones. Rat small intestine was used for this purpose (Figure 2.5).

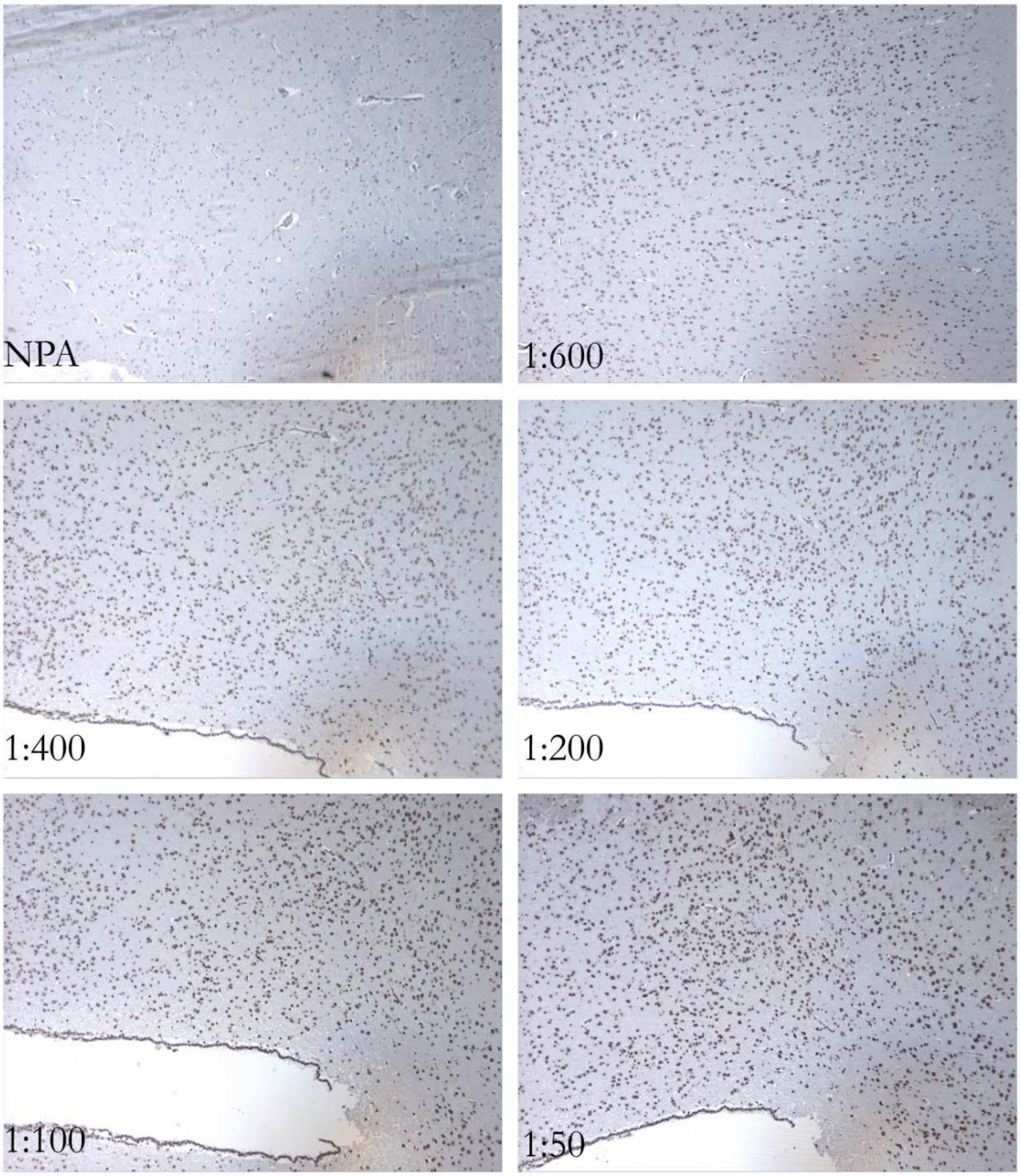


Figure 2.4 Optimisation of primary antibody dilution – 5MC.
4 x Magnification light microscope images of rat brain showing the third ventricle (3V) and associated hypothalamic nuclei. Primary antibody dilutions for 5MC increase from No Primary Antibody (NPA) to 1:50. 1:200 was selected as the optimal dilution of 5MC as it exhibited the strongest nuclear signal with the least background.

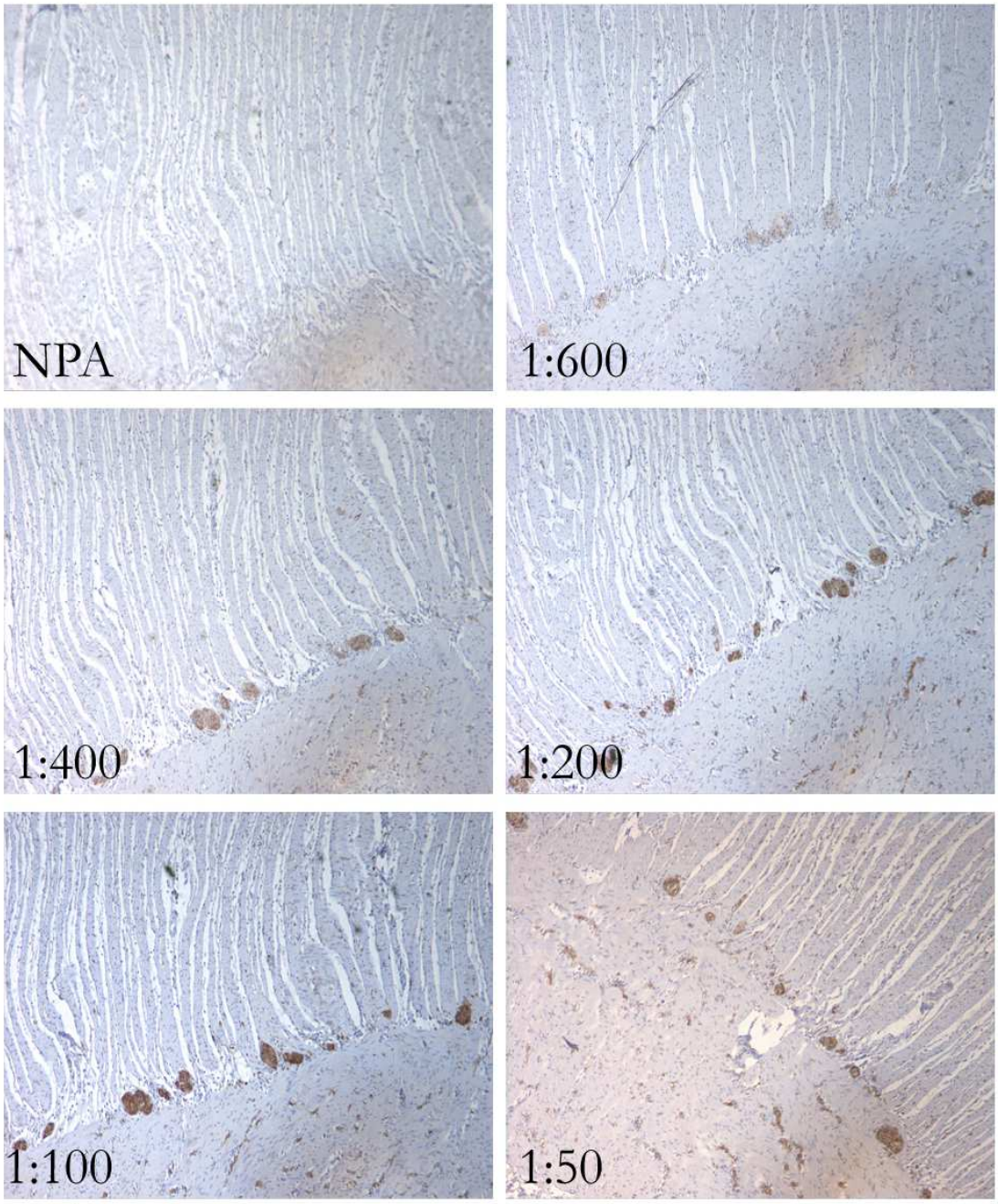


Figure 2.5 Optimisation of primary antibody dilution – NSE.
4 x Magnification light microscope images of rat small intestine showing villi and Aurebach's plexus. Primary antibody dilutions for NSE increase from No Primary Antibody (NPA) to 1:50. 1:200 was selected as the optimal dilution of NSE as it exhibited the strongest nuclear signal with the least background.

2.3.6 Secondary antibody selection

Fluorescence conjugated secondary antibodies were selected appropriate to the isotype of the primary antibody. Rho Red (IgG1) was selected for 5MC, while FITC (IgG2b) was selected for NSE and B3T (Invitrogen, Paisley, UK).

2.3.7 Fluorescent secondary antibodies

The optimal primary antibody binding conditions were carried over from the chromophore optimisation as described above. Following incubation with the primary antibody the slides were twice washed with PBS before being incubated with fluorescent secondary appropriate to isotype as described above. Slides were incubated in the dark for onr hour at room temperature before being washed twice with PBS. Slides were then dried and coverslips mounted with Vectorshield mountant (Vector, Peterborough, UK).

2.3.8 Imaging

Images of the slides were captured using an Axio Imager A2 (Carl Zeiss, Welwyn Garden City, UK). Images were captured at 20x, 40x and 63x and the three magnifications compared statistically for sensitivity and coverage. Examples of magnification power in brain tissue are shown in Figure 2.6. Fluorescent conjugates were excited at 540nm (Rho Red) and 495nm (FITC) and the images stacked to create a composite (Figure 2.7). For the purposes of fluorescent quantification the exposure time for the 5MC secondary antibody was fixed at 100ms for all images. Exposure time for B3T was determined automatically as this did not affect the quantification process.

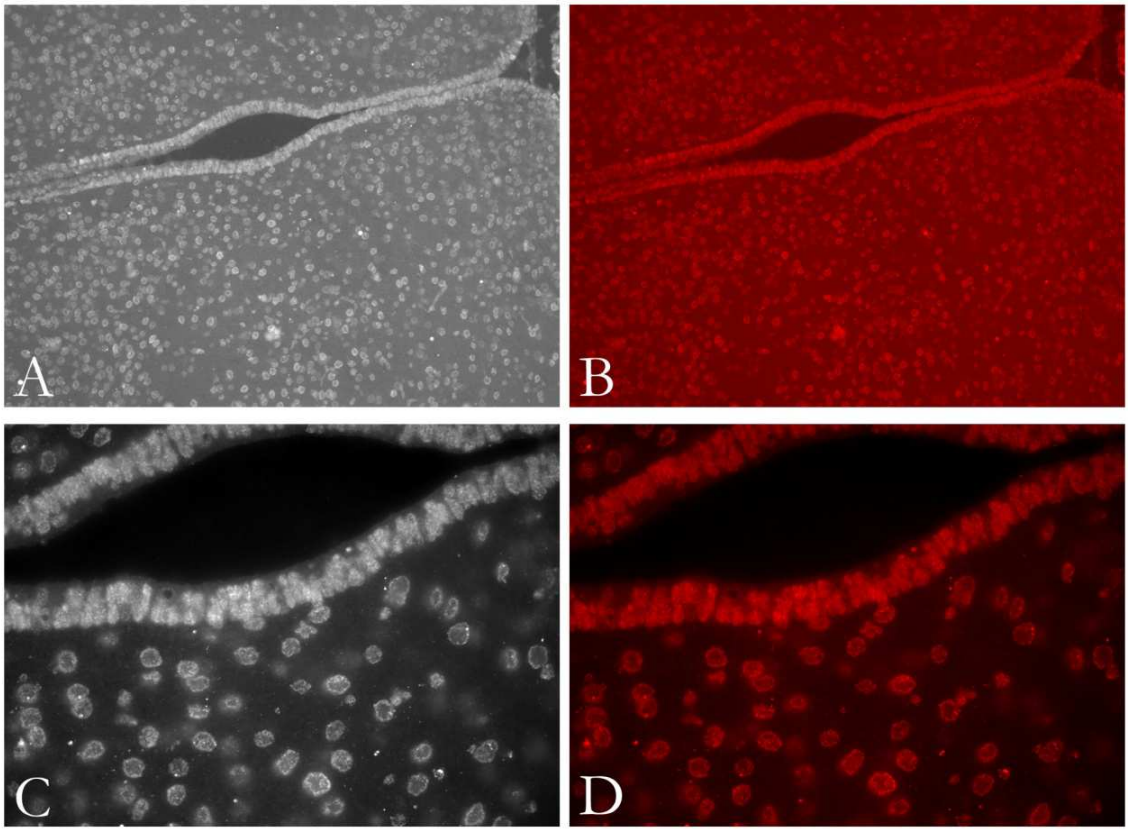


Figure 2.6 5MC Fluorescent microscopy of hypothalamus at 20 x and 63x magnification.
Fluorescent microscope images of rat brain showing third ventricle (3V) and associated hypothamalmic nuclei. A - 20x magnification eight bit non coloured image, B - 20x magnification RGB psuedo red coloured image, C - 63x magnification eight bit non coloured image, D - 63x magnification RGB psuedo red coloured image.

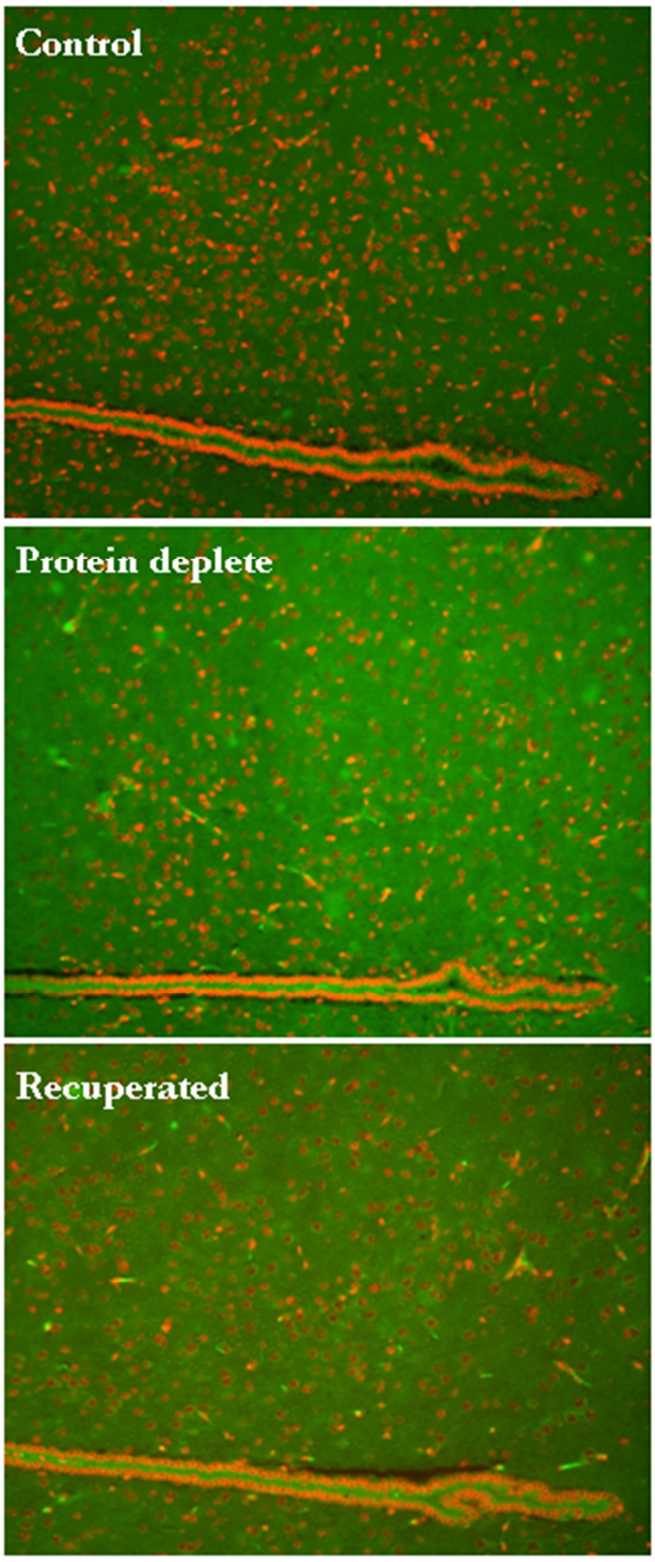


Figure 2.7 Fluorescent double labelling of hypothalamic sections using 5MC (Red) and B3T (Green) taken at 20x magnification.

Fluorescent microscope images showing third ventricle (3V) of the hypothalamus. Nuclei stained with false red 5MC antibody, neuronal cells stained with false green B3T marker.

2.3.9 Quantification of fluorescence

Fluorescence was quantified using the ImageJ 1.43 software package (NIH, Maryland, US). Full colour images were converted to 8-bit file format and nuclei counted using the Particle Analysis/Nucleus Counter plugin. DNA methylation was quantified by measuring fluorescence intensity in the nuclei using the IntesityCount utility. Mean methylation was calculated across the nuclei in a 500 x 500 pixel region of interest corresponding with the paraventricular nucleus. Images were captured on both the left sides of the hypothalamus for each section. Each section was imaged in duplicate resulting in four fluorescence measurements for each individual animal.

Images were captured for a total of 18 animals, six from the control group, seven from the protein restricted group and five from the recuperated group. Each slide included a no primary antibody section (NPA) as a control.

Images were assessed using ImageJ Nucleus Counter. Nucleus particle size was set between 100 and 2500. Fluorescence intensity in these nuclei was then quantified using IntensityCount (Figure 2.8).

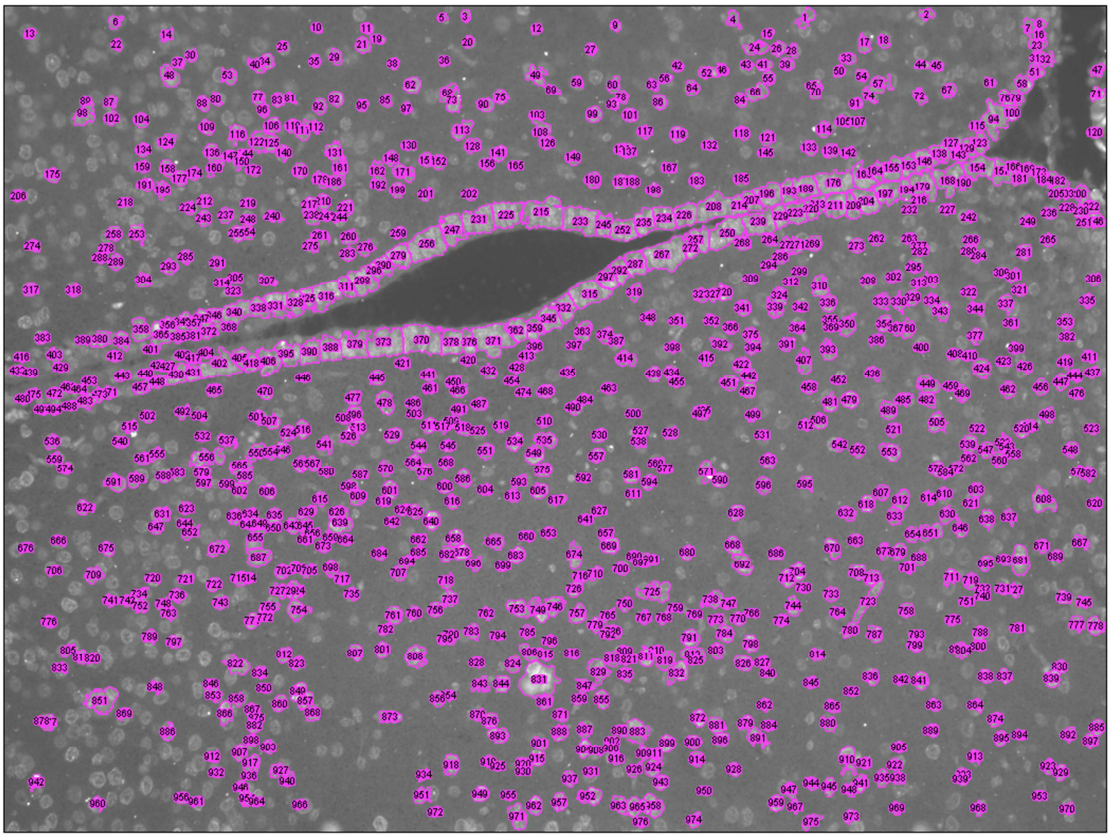


Figure 2.8 Hypothalamic section with ImageJ nucleus counter overlay.
Purple highlighted and numbered sections represent regions of interest (ROIs) as generated by Nuclear Counter. Intensity in each one of these numbered sections was individually quantified.

Data analysis

Fluorescence intensities across magnification powers were compared using Wilcoxon signed-rank test. The correlation between repeats within the same section was assessed using Spearman's correlation. The mean fluorescence intensities for all three exposure groups were compared using the Kruskal-Wallis test.

2.4 Results

2.4.1 Selection of optimum magnification for fluorescence quantification

Optimum magnification was determined by serial imaging of identical sections at three different magnifications, x20, x40 and x63. Six samples were imaged, two from each intervention group in order to compare sensitivity of the three magnifications. Table 2.1 presents Median and inter-quartile ranges (IQR) for each magnification. There were no significant differences between any of the groups compared at any of the three magnifications shown by Wilcoxon signed-rank test (Table 2.1). As a result of this analysis x20 was selected as the optimum magnification because it enabled a greater number of nuclei within the field of view to be imaged with no significant reduction in sensitivity. For instance across the six samples represented in Table 2.1, the average number of nuclei captured at x20, x40, x63 magnifications were 140, 100 and 55 respectively.

Magnification	n	Median (IQR) Intensity*	Group comparison	z‡	p-value
x20	6	1024.80 (797.51, 1239.71)	20 vs. 40	-0.105	0.916
x40	6	1029.83 (970.59, 1121.47)	40 vs. 63	-0.315	0.752
x63	6	1114.09 (797.92, 1263.77)	20 vs. 63	-0.524	0.600

Table 2.1 Comparison of fluorescence intensities across magnification groups.
*Fluorescence intensities standardized for exposure time (i.e. intensity per second); ‡Group comparison by Wilcoxon signed-rank test; IQR – Interquartile range

2.4.2 Quantification and analysis of fluorescence across intervention groups

A total of 18 animals across all three intervention groups were imaged and quantified using the optimized protocol described above. Two serial sections from each animal were analysed with two fluorescence measurements per section to assess inter-section variability. The correlation in fluorescence between repeats was assessed using a Spearman's correlation test. As shown graphically in Figure 2.9, in the control group fluorescence measurements between repeats were highly positively correlated and modestly correlated across sections. However in both the protein and recuperated groups repeat measurements were highly positively correlated whereas inter-section measurements showed strong negative correlations, particularly in the protein group. Consequently subsequent association analyses were performed using within-section means for each of the two sections separately.

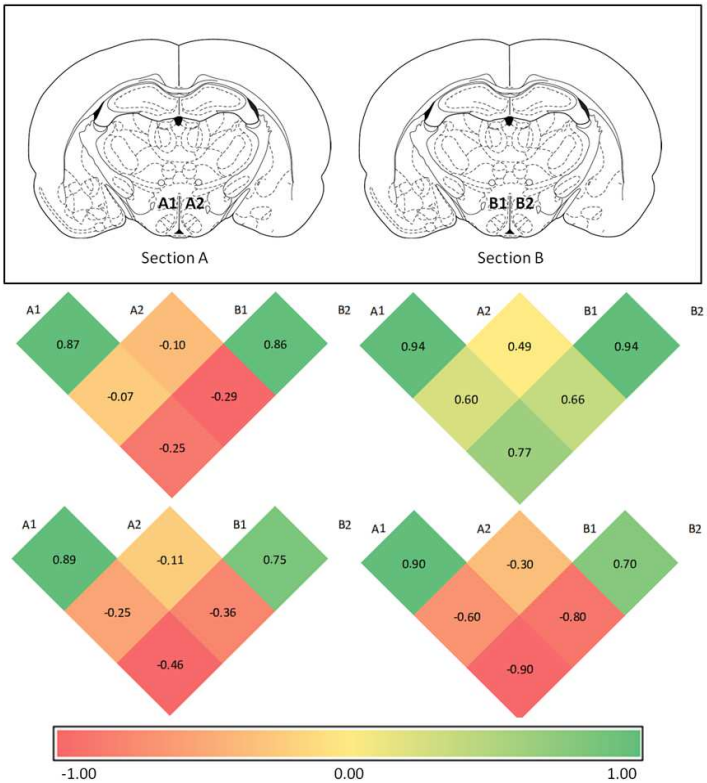


Figure 2.9 Correlation between repeat measurements in all samples within each intervention group.

The upper section depicts the two serial sections with repeat measurement locations numbered one to four; with one and two located on section A, and three and four on section B. The lower section depicts pairwise correlation in fluorescence intensities between these repeats; numbers within coloured boxes are Spearman's correlation Rho values.

Table 2.2 shows median and IQR for the fluorescence intensities across the intervention groups. No significant differences between groups were observed in either section A or section B when tested by Kruskal-Wallis (Table 2.2).

		All groups		Kruskal-Wallis	
Fluorescence site	Group	n	Median (IQR) Intensity	χ^2	p-value
A1	All	18	102.51 (67.26, 120.30)		
	Control	6	101.50 (34.24, 114.78)		
	Protein	7	92.67 (60.02, 133.04)		
	Recuperated	5	112.34 (90.71, 120.30)		
A2	All	18	99.22 (47.52, 122.48)		
	Control	6	96.64 (42.48, 115.27)		
	Protein	7	99.48 (68.62, 136.54)		
	Recuperated	5	118.40 (34.43, 131.08)		
B1	All	18	91.75 (43.55, 100.82)		
	Control	6	92.98 (52.87, 100.82)		
	Protein	7	90.10 (42.60, 110.61)		
	Recuperated	5	91.84 (91.66, 92.20)		
B2	All	18	80.36 (44.32, 102.59)		
	Control	6	97.30 (47.48, 114.67)		
	Protein	7	51.66 (43.84, 96.16)		
	Recuperated	5	71.96 (53.40, 88.76)		
Mean Section A	All	18	99.07 (56.53, 123.98)		
	Control	6	99.07 (37.67, 115.03)		
	Protein	7	89.93 (74.82, 134.79)	0.94	0.624
	Recuperated	5	115.37 (56.53, 125.69)		
Mean Section B	All	18	85.98 (47.13, 103.39)		
	Control	6	96.64 (50.18, 104.75)		
	Protein	7	68.82 (43.94, 103.39)	0.91	0.634
	Recuperated	5	82.08 (72.62, 90.21)		

Table 2.2 Fluorescence intensity compared between the three intervention groups.
IQR – Interquartile range. Kruskal Wallis Chi squared and p values shown for differences between three exposure groups in mean sections only.

The fluorescence intensities across section repeats were compared by creating scatter plots of the repeats within each section against each other (Figure 2.10). Within all six sections there was a positive association showing good reproducibility.

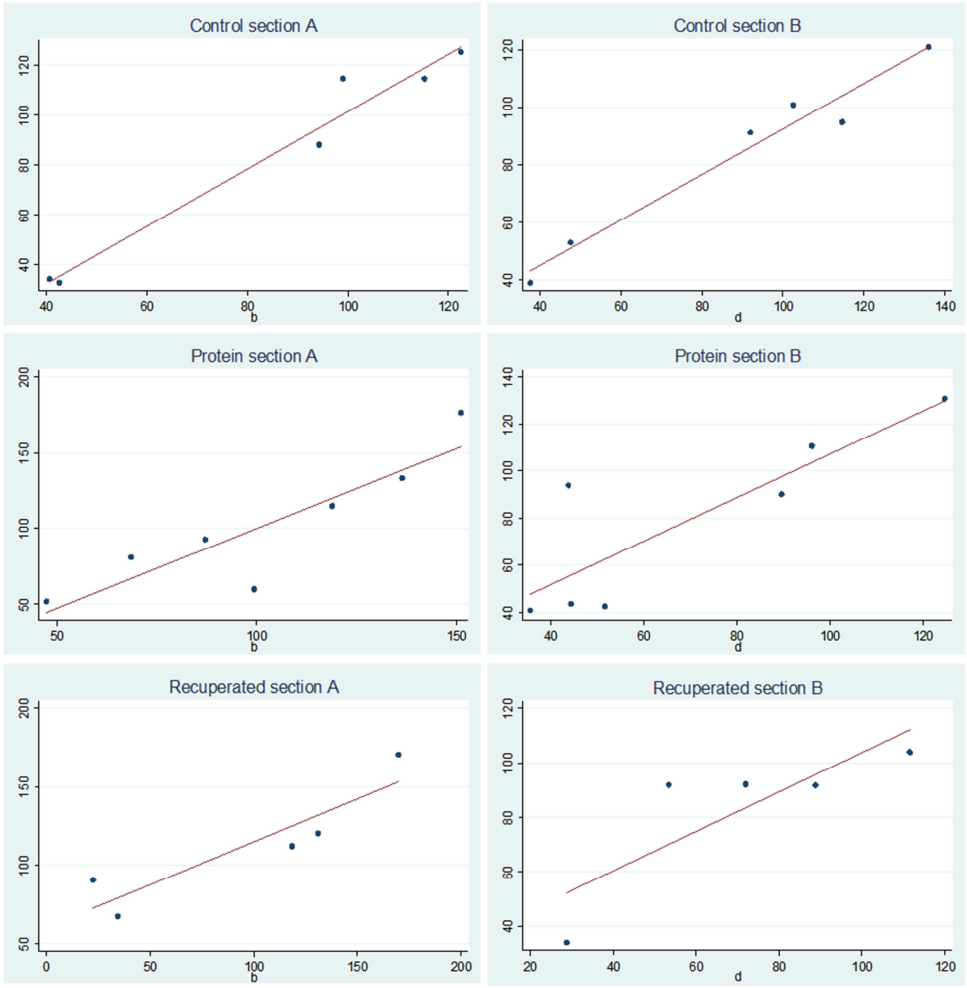


Figure 2.10 Scatter plots for fluorescence intensities across within section repeats by intervention group.

The individual exposure groups were also compared by creating scatter plots of the mean fluorescence intensities for each repeat against each other. There was a positive association between the fluorescence intensities in the control group but negative associations between the fluorescence intensities in the protein and recuperated groups (Figure 2.11).

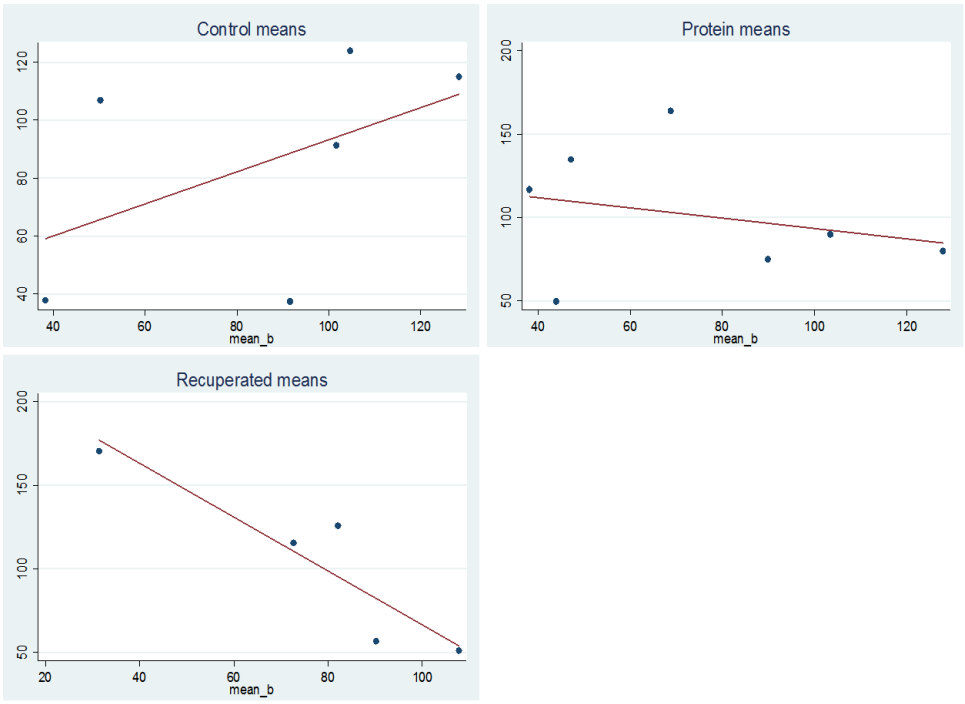


Figure 2.11 Scatter plots for fluorescence intensities across section means by intervention group.

The unusual observation that there was negative association between fluorescence intensity in the protein and recuperated groups lead us to re-plot these data as a scatter plot to compare the distribution of fluorescence intensities between intervention group, stratified by section (Figure 2.12). The scatter plots show that there is a greater level of variation in fluorescence intensity in section B compared to section A, apart from the recuperated group. This would seem to be the source of the poor association between the fluorescence intensities in the case of the recuperated group.

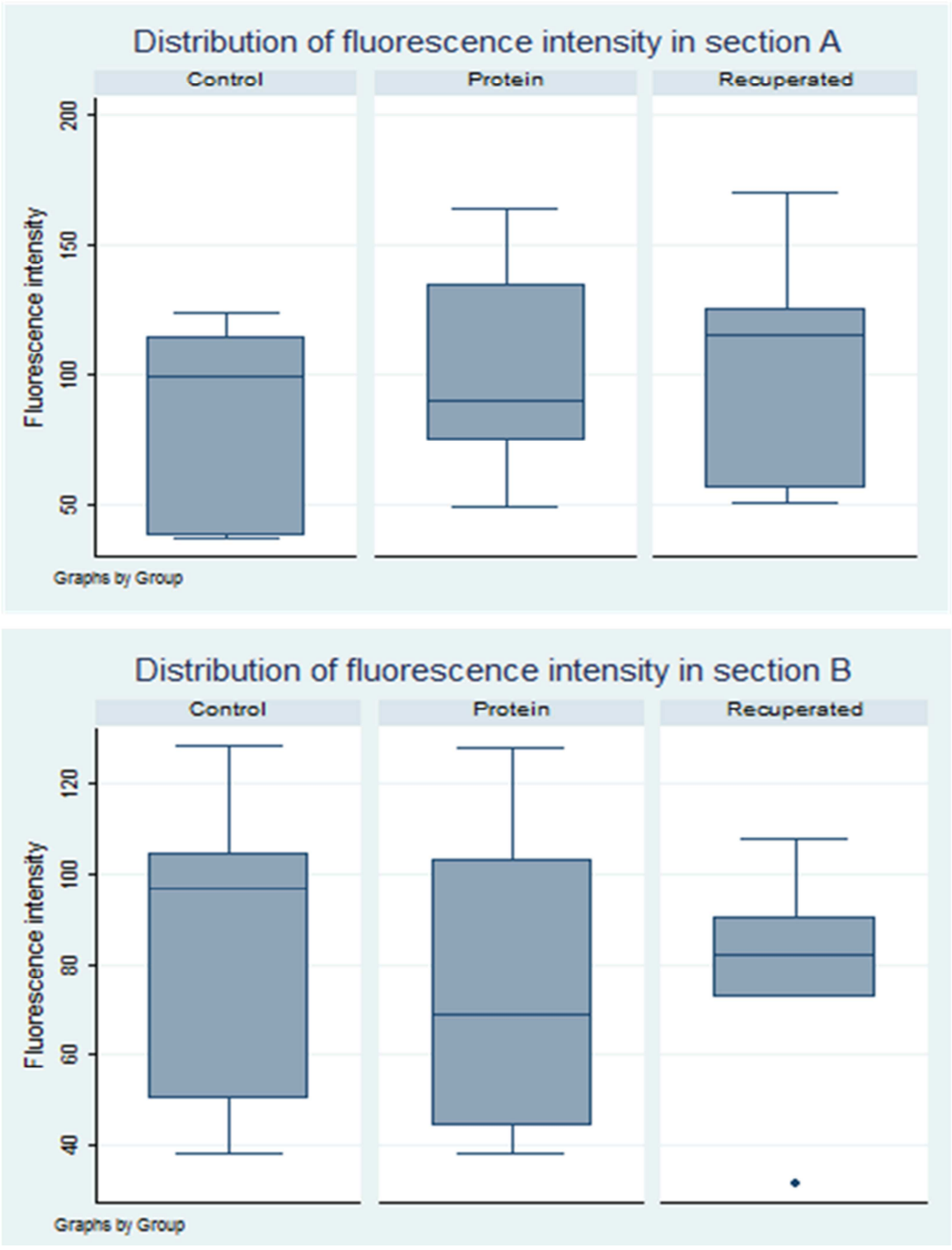


Figure 2.12 Distribution of fluorescence intensities across all three intervention groups, sections A and B compared.

2.5 Discussion

The techniques and analysis methods for the immunohistochemical (IHC) staining of DNA methylation in brain tissue described in this chapter had not been applied previously and were developed and optimised as part of this study. Hence, a considerable amount of technical and methodological development was required before meaningful data could be obtained.

Brain tissue was obtained from a rodent model. Samples were initially available as Paraformaldehyde-perfused, frozen sections and the use of these sections in IHC was explored. However as a new tranche of rodents were being culled and their brains removed it was decided that in order to best preserve the brain architecture and prevent freezing artefacts the best fixation method was formalin fixation of whole brain, followed by paraffin embedding. Perfused sections are seen by some as preferable for these kinds of experiments as the perfusion process removes red blood cells and can prevent tissue damage (Scouten, 2010). However in practise the paraffin embedded sections proved suitable for this application following careful optimisation.

The choice of primary antibodies both to a marker of methylation and a known neuronal marker was an important one. In the case of the methylation mark a number of antibodies are commonly used for this type of work. The DNA methyltransferase (DNMT) family of enzymes (DNMT 1, 2 and 3) work to catalyse the addition of a methyl group to the DNA and levels of DNMT's are a surrogate for levels of DNA methylation. The Methyl-CpG binding proteins (MBD1-4) are also candidates as they bind directly to the methyl group of the DNA and show localisation. The addition of a hydroxymethyl group to the DNA has been of particular interest to a number of groups recently (Wossidlo et al., 2011) and as such antibodies to 5hMC were considered. The antibody that was eventually selected however was one raised to 5-methylcytosine for the reason that this provides a direct marker of DNA methylation within the nucleus. This helped to inform the selection of a neuronal cellular marker as the localisation of 5MC to the nucleus meant that it would be best practise to select a cellular marker localised to the cytosol. There are a number of neuronal cellular markers available however the most commonly used are neurone specific enolase and Beta Tubulin (Subtype III). Both of these markers are specific to neuronal cells and localise to the cytoplasm and so represented ideal antibody choices for our application.

A key decision relating to the quantification of DNA methylation levels within hypothalamic nuclei was defining the optimal magnification to use. The greater the magnification used on the fluorescence microscope the more detail discernable in individual nuclei. There exists a trade off however with the number of nuclei present within the field of view. As shown in the text as the magnification was increased from 20x to 63x the number of nuclei available to quantify drops from 140 to 55. As such a comparison between all three magnifications made it possible to discern whether higher magnifications allow more accurate quantification of DNA methylation. Table 2.1 demonstrates that there was no significant difference between the fluorescence intensities measured across the three magnifications. This allowed a larger number of nuclei to be assayed, giving a more accurate overall picture of DNA methylation within the hypothalamus without any significant loss of sensitivity.

In order to test the reproducibility of fluorescence measurements made on each section, repeats were taken on each section as shown in Figure 2.7. Measurements taken from repeats on the same section were shown to be highly correlated in control animals but significantly less so in the protein and recuperated groups. The within section fluorescence measurements were however highly correlated. Thus intra-section reproducibility was much greater that inter-section reproducibility despite sections being mounted on the same slide and stained at the same time. The root of these differences between sections may be in the relatively small sample size for each group as well as the overall lower levels of fluorescence in section B versus section A, noticeable particularly in the recuperated group (82.08 fluorescence units (FU) vs. 115.37 FU).

No differences were observed between fluorescence intensities between the three treatment groups analysed. There are a number of possible explanations. Firstly, this technique might not possess a high enough sensitivity to detect subtle alterations in DNA methylation. Further, IHC using an antibody specific to 5MC is limited to providing a global overview of the 5-methyl cytosine content of each nucleus. It cannot discern if there are gene specific changes in DNA methylation and indeed if there are they could be in differing directions, which would serve to mask each other.

It is possible that no changes in DNA methylation were detected because there were no changes to the levels of DNA methylation in the hypothalamic region of animals in this model in response to dietary insult. However this is unlikely given that there is significant evidence from other models that hypothalamic gene expression levels are altered significantly in response to insult (Bouret et al., 2008, Kirk et al., 2009). There is also a

growing body of evidence to suggest that DNA methylation varies significantly in brain tissues (Zhang et al., 2010a).

Given the lack of any association between treatment and 5-MC measured using IHC in these experiments, there are a number of additional IHC-based strategies which could be explored to further investigate the role of DNA methylation in the programming of phenotype. The IHC techniques described here could be optimised with antibodies specific to a range of other molecules that associate with 5-methylcytosine. One such example is the DNA methyltransferase family (DNMT) (Lin et al., 2007) which is involved in the maintenance of DNA methylation throughout the genome. Additionally other epigenetic-related proteins could serve as targets for IHC including activation-induced cytidine deaminase (AID) (Larijani et al., 2005), the methyl binding domain family (MBD) (Fujita et al., 2003) and methyl-CpG-binding domain protein 2 (MeCP2) (Fuks et al., 2003). There is also burgeoning interest in 5-hydroxymethylcytosine (5-hMC), the addition of both a methyl and a hydroxy group to cytosine residues in the genome (Iqbal et al., 2011), and this may be of particular interest given that 5-hMC residues have been found in brain tissues (Kriaucionis and Heintz, 2009).

As an alternative to IHC analysis of tissues it would also be plausible to apply more molecular based techniques to try and discern any difference in DNA methylation between the brain tissues of exposure groups. For example laser capture microdissection could be utilised (Gagnon et al., 2010, Eberle et al., 2010) to isolate specific brain regions from pre cut sections. DNA extraction could then be performed on the relevant areas and downstream analyses such as MeDIP-chip (Methyl-DNA immunoprecipitation followed by tiling array), which would allow DNA methylation to be mapped and measured across the genome (Palmke et al., 2011). Alternatively conventional dissection techniques could be used and DNA and RNA extracted to allow gene specific DNA methylation and gene expression experiments to be performed (Gibbs et al., 2010).

This area could also be further investigated by applying other IHC or molecular techniques mentioned above on an array of different brain areas. Although the paraventricular nucleus has been well characterised as a centre of appetite control (Kalra et al., 1991) there are a wealth of other regions which may be of interest (Berthoud and Morrison, 2008). These include the caudal brain stem which has been implicated in the signal transduction as part of the leptin signalling pathway (Grill et al., 2002) and the corticolimbic system (Berthoud, 2004) which governs both emotional and cognitive responses to food and hunger. Although the brain plays a key role in the control of appetite and therefore energy balance

(Richard et al., 2009) it would be interesting to assess tissue specific DNA methylation in other tissue types in which DNA methylation has been implicated in the pathogenesis of obesity including liver (Gomez-Acevedo et al., 2011) and white adipose tissue (Pinnick and Karpe, 2010).

It is also important to consider that in relative terms the nutritional insult to which the experimental animals were exposed is a mild one. Further work could utilise a more extreme nutritional exposure in order to evoke a stronger response, such as a 'junk food' (Ong and Muhlhausler, 2011) or high fat diet (Vucetic et al., 2010) or alternatively in an extreme phenotype model such as the Zucker rat (Williams and Schalinske, 2011) or a disease based model such as diabetes (Nieman and Schalinske, 2011).

In summary, although these data show no significant differences between the three exposure groups, the area of hypothalamic programming of obesity is still of interest and may benefit from a candidate driven, gene specific approach to the investigation of DNA methylation (Wang et al., 2010a).

Chapter 3: Developing *In Silico* Approaches for Target Gene Selection and Quantitative DNA Methylation Analysis

3.1 Background

In recent years, the quantum leap in development of methodology for investigating the genome using both sequencing and microarray approaches has brought with it both huge advantages and a number of potential pitfalls (Khatri and Ghici, 2005). It is now possible to perform multiplex expression experiments on many thousands of loci in a matter of hours. The rate-limiting steps in the production of high quality data from these experiments are now in the timely and efficient analysis of array or sequence data, in the interpretation of the resulting observations and in generation of subsequent hypotheses. The size of a gene expression microarray dataset makes these steps both a significant intellectual and a significant bioinformatics challenge. Our goal when assessing a gene expression dataset is to assign functional meaning to any changes observed. Traditional approaches filter the output dataset based on criteria intended to quantify the size and significance of any change in expression at a gene specific level (Smirnov et al., 2009). These criteria include fold change, p value and, often, minimum levels of expression. Such approaches advocate the selection of a top hits list based on cut offs – be they a fold change of greater than 2.0 or a p value of <0.05 (Haroon et al., 2010). Although this approach is a reasonable one and leads to the selection of robust functional targets, the application of somewhat arbitrary cut off points may exclude a large amount of the dataset and so the loss of potentially interesting information.

Having identified genes that are differentially expressed in the two (or more) conditions under investigation, a further commonly used step is text data mining. Text data mining (TDM) is the extraction of information of interest from published documents (Krallinger et al., 2008). This information usually relates to a specific gene or protein and its association with a disease or other biological function. Electronic resources are now the first port of call for any researcher looking to discern links between genes and outcomes of interest and indeed are key in hypothesis generation. As a result text data mining is becoming a increasingly utilised research technique (Zvi et al., 2008). A key question when considering TDM approaches is the level of robustness of the text analysis. TDM is used typically to

assess disease candidates and protein-protein interactions (Chen et al., 2009b, Köhler et al., 2008). However there are also several tools available for assessing links between a gene of interest and the potential for differential methylation (Krallinger et al., 2010). The robustness of TDM approaches is admittedly variable because of the plethora of different methods, however what all methods have in common is the possibility of 'losing' some output due to problems with search string creation or wording issues. For example if we were to undertake a text data search for the gene name *IGF2R* we would also want information on IGFIIR and any other forms of the phrase to be included. This is a key step in the design of a TDM program. The robustness of TDM in the context of the biomedical field is also largely dependent on which databases (MedLine, PubMed etc.) are polled. Numerous tools are available to assess these literature resources. MeInfo Text (MIT) (Fang et al., 2008) is one such tool. MIT text data mines a number of sources including NCBI Entrez for gene information, HPRD and IntAct for protein interactions, HPRD and KEGG for pathway types and PubMed for text strings. Although designed for use in the study of DNA methylation in cancer, MIT provides a useful tool alongside traditional non-automated text data mining techniques for investigation of genes whose promoters may be differentially methylated in particular circumstances.

More recently, novel approaches to analysing large gene list based datasets have been developed (Zhong et al., 2010). Whereas previous techniques selected candidates based purely on the magnitude of expression, newer methods also incorporate gene function. Published functional data on all genes showing significant expression changes in a given dataset (for example a gene expression array) is assessed and potential biological pathways of interest are generated. This allows the selection of candidates which are not only differentially expressed but are also relevant in terms of a specific pathway or disease of interest.

There is an increasing body of evidence that gene expression is regulated by, and correlates inversely with, the methylation status of that gene's promoter (Thompson et al., 2010b, Sears et al., 2011). High levels of gene expression do not always correlate with low levels of DNA methylation (Bell et al., 2011) (Figure 3.1) which could have implications on the present study given that it is an attempt to show that aberrant DNA methylation affects gene expression. The majority of genes will however follow the usual convention and show low levels of DNA methylation when genes are highly expressed. Gene expression datasets

represent a useful starting point in the identification of epigenetically regulated target genes. This approach allows a complete gene expression dataset to be analysed functionally using curated data rather than relying purely on gene expression levels to inform the genes which may be of further interest.

An advantage to this approach is that it represents a targeted and logical way to prioritise candidates from a large gene list. Using experimentally validated functional data on genes within the gene list to inform our selection of candidates make it more likely that the genes taken forward for further analysis will be those that are of interest in the context of our chosen outcome, in this case obesity and related sequelae. Each assay taken forward into the laboratory requires optimisation and validation that is costly both in terms of time and money. As such a robust approach to target identification allows only those genes that are most relevant to progress through to the 'wet laboratory' phase.

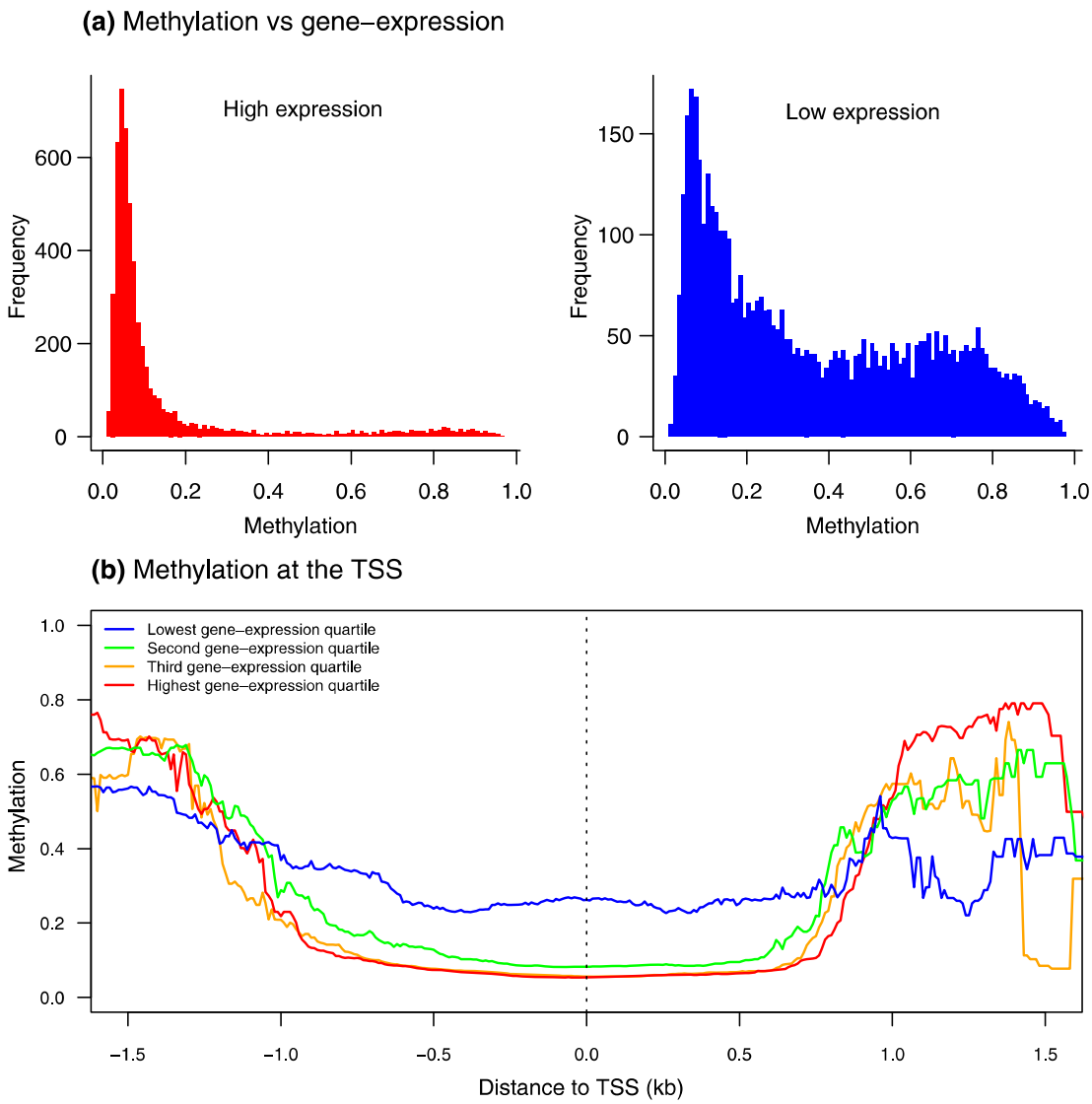


Figure 3.1 DNA methylation is negatively correlated with gene expression.
(a) Methylation levels are low in the top quartile of highly expressed genes (left), and high in the bottom quartile of lowly expressed genes (right), looking across 12,670 autosomal genes. (b) Methylation levels with respect to the TSS in sets of genes categorized by gene expression levels, from highest (red) to lowest (blue), using the quartiles of gene expression with respect to gene expression means, where fitted lines represent running median levels. (Bell et al., 2011).

3.2 Aims and objectives

The aims of this chapter were i) to develop a bioinformatic pipeline for the analysis of expression data obtained from microarray analysis and ii) to use this pipeline to identify obesity-related genes that were differentially expressed in the offspring of female mice (dams) fed an obesogenic diet where the differential expression could be due to altered methylation of the gene promoters. The pipeline was used to interrogate a gene expression array dataset to produce a short-list of candidates for further investigation of promoter methylation using pyroseqeuncing. The final aim of the work was to design pyrosequencer-based assays for quantifying methylation of specific CpG sites within the promoters of the short-listed genes.

3.3 Methods

An initial gene list of over 19,000 was interrogated using a step-wise workflow with attrition at each stage facilitating the generation of a practical number of top hits to take forward into downstream analyses (Figure 3.2).

3.3.1 Mouse gene expression dataset

The gene expression dataset utilised for this work was kindly provided by Professor Lucilla Poston of King's College London (Samuelsson et al., 2008). This dataset was derived from analysis of RNA from tissues from the female offspring of C57BL/6J mice. The mouse dams were randomised onto either a standard mouse chow (7% simple sugars, 3% fat, 50% polysaccharide, 15% protein, energy 3.5 kcal/g, n=20), or a highly palatable obesogenic diet (10% simple sugars, 20% animal lard 28% polysaccharide, 23% protein, energy 4.5 kcal/g, n=30) and fed ad libitum. After six weeks on the appropriate diet, the animals were mated and stayed on the obesogenic diet throughout gestation. Tissues were collected from the offspring both at birth and at six weeks of age. Total RNA was extracted from the liver, heart and white adipose tissue (WAT). Sufficient RNA was available from liver and heart only in new born pups due to the lack of a substantial fat pad for RNA extraction (Table 3.1) (Samuelsson et al., 2008) A total. Total RNA was hybridised to Illumina Sentrix® MouseRef-8 Expression BeadChips with n=8 for each diet giving a total of n=32. Raw expression values for each tissue and time point were sorted

by p-value and all those with a between diets p-value greater than 0.05 were excluded from the analysis.

	Foetal	Neonatal
WAT	✗	✓
Liver	✓	✓
Heart	✓	✓

Table 3.1 Tissue availability at both time points.
No white adipose tissue (WAT) was available at the foetal timepoint due to insufficient fat pad size.

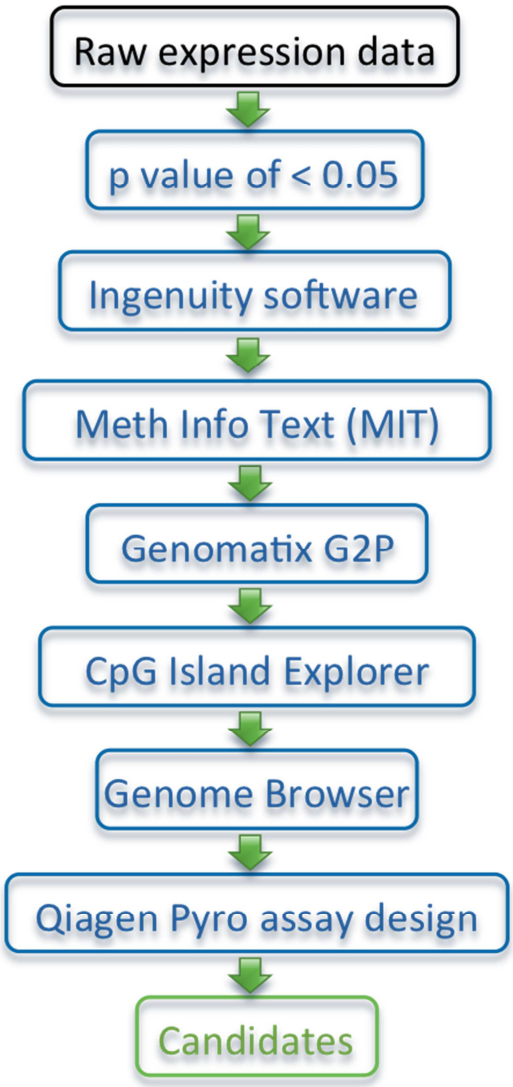


Figure 3.2 Target gene selection workflow
Text within the blue boxes shows each processing method.

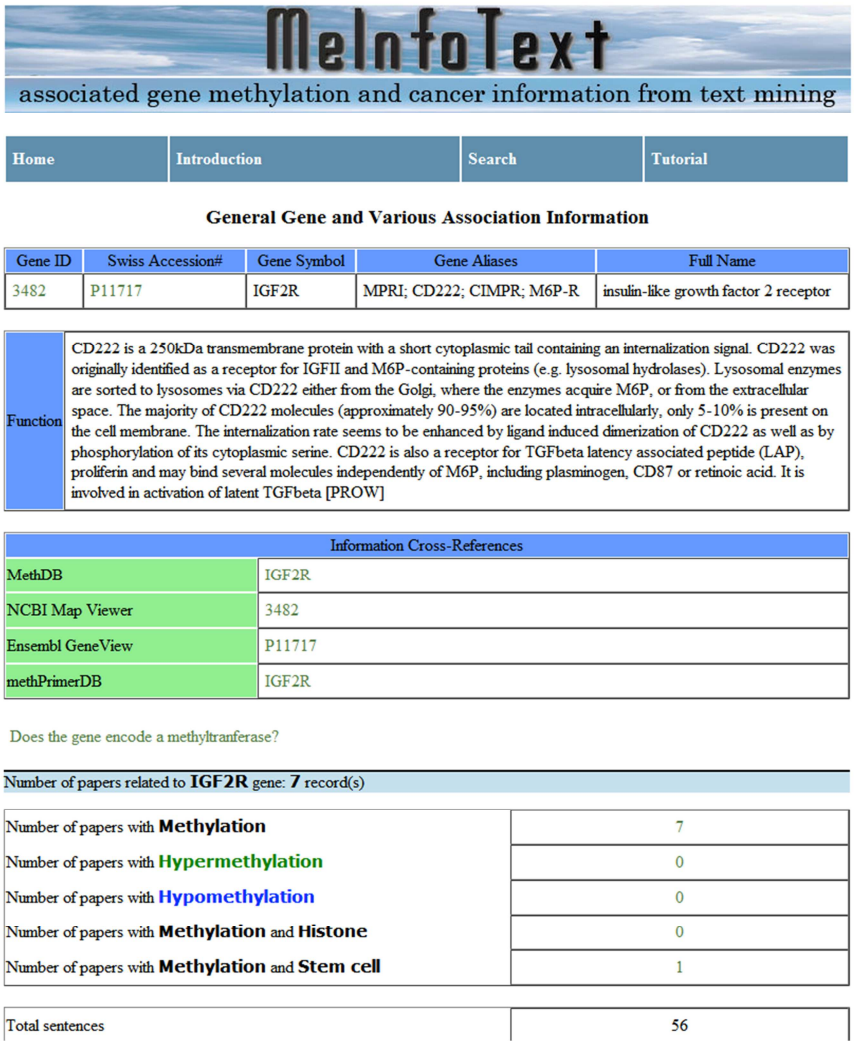
3.3.2 Pathway Analysis

The complete 'raw' gene lists described above were uploaded to Ingenuity Systems Inc.'s IPA pathway analysis software (Myslobodsky, 2008). IPA utilises a proprietary dataset known as the 'Knowledge Base', a manually reviewed and curated database of interactions between biological elements. IPA models the molecular interactions between thousands of genes, RNAs, proteins, cells, drugs and diseases. This information was abstracted from peer-reviewed scientific publications and from commonly used tools such as National Centre for Biotechnology Information (NCBI) Reference Sequence (RefSeq), Online Mendelian Inheritance in Man (OMIM), Kyoto Encyclopaedia of Genes and Genomes (KEGG) and Genome-wide Association Studies (GWAS). The result is a database of approximately 200,000 scientific articles as well as information on approximately 10,000 human, 8,000 mouse and 5000 rat genes. The software incorporates existing knowledge on the molecular interactions between all of these elements allowing the 'Knowledge Base' to be polled for a large variety of biological questions. The database structure is defined by a strict ontology scheme which allows the great majority of molecules to be referenced against the 'Knowledge Base' and placed into a biologically relevant pathway (Calvano et al., 2005).

In practical terms, IPA is a suite of software tools that enable the interrogation of complex datasets including gene expression arrays. There were a total of two time points and three tissue types from which RNA and therefore expression data was available (Table 3.1). As such there were effectively five separate gene lists that were compared to the IPA database. These five gene lists were first mapped to IPA's own gene ontology (GO) list to ensure correct functional characterisation of the genes. This allowed the gene lists to be reliably compared with the IPA 'interactome', a functional database of molecular interactions. Once the five gene lists were satisfactorily mapped to the GO structure, relevant queries were built to test the biological relevance of any pathway wide changes in the gene expression datasets. Obesity was the main pathology of interest, but as it is linked to a number of co-morbidities these were also included in the analysis. As such the following five pathways were selected: obesity, cardiovascular disease, leptin signalling, insulin signalling and type 2 diabetes. DNA methylation was selected to supplement the metabolic pathways in light of our group's interest in epigenetic mechanisms in the early life programming of obesity risk.

3.3.3 Text data mining for evidence of promoter methylation

The gene list was analysed using text data mining to ascertain if any of the candidates had been identified previously as differentially methylated in any circumstance. This data mining was performed using the freely available tool MIT which is described briefly above. Output genes from IPA that showed associations with DNA methylation progressed to the next stage of the candidate selection process (Figure 3.3).


MeInfoText

associated gene methylation and cancer information from text mining

Home | Introduction | Search | Tutorial

General Gene and Various Association Information

Gene ID	Swiss Accession#	Gene Symbol	Gene Aliases	Full Name
3482	P11717	IGF2R	MPRI; CD222; CIMPR; M6P-R	insulin-like growth factor 2 receptor

Function	CD222 is a 250kDa transmembrane protein with a short cytoplasmic tail containing an internalization signal. CD222 was originally identified as a receptor for IGFII and M6P-containing proteins (e.g. lysosomal hydrolases). Lysosomal enzymes are sorted to lysosomes via CD222 either from the Golgi, where the enzymes acquire M6P, or from the extracellular space. The majority of CD222 molecules (approximately 90-95%) are located intracellularly, only 5-10% is present on the cell membrane. The internalization rate seems to be enhanced by ligand induced dimerization of CD222 as well as by phosphorylation of its cytoplasmic serine. CD222 is also a receptor for TGFbeta latency associated peptide (LAP), proliferin and may bind several molecules independently of M6P, including plasminogen, CD87 or retinoic acid. It is involved in activation of latent TGFbeta [PROW]

Information Cross-References	
MethDB	IGF2R
NCBI Map Viewer	3482
Ensembl GeneView	P11717
methPrimerDB	IGF2R

Does the gene encode a methyltranferase?

Number of papers related to **IGF2R** gene: **7** record(s)

Number of papers with **Methylation**	7
Number of papers with **Hypermethylation**	0
Number of papers with **Hypomethylation**	0
Number of papers with **Methylation** and **Histone**	0
Number of papers with **Methylation** and **Stem cell**	1

Total sentences	56

Figure 3.3 Screen capture from MIT application showing text data mining result.
The gene queried (*IGF2R*) is shown mapped to NCBI, Ensembl and methPrimer. The number of papers published showing the query gene associated with methylation, hypomethylation and hypermethylation.

3.3.4 Promoter searching

Promoter regions are key regulatory elements found 5' to the coding regions of genes. They play a key role in regulation of expression of the corresponding gene and, as mentioned above may be subject to control by differential methylation in certain genes. To assess whether any of the candidate genes produced from the IPA analysis are subject to regulation via differential methylation, the next step was to search for the presence of putative methylated regions in their promoters. Annotated and experimentally verified promoter sequences for the queried genes were generated by the Genome2Promoter utility provided as part of the Genomatix software suite (Genomatix Ltd, London, UK). The output from this analysis includes the chromosomal location as well as a description of the query gene alongside the promoter sequences. The extracted promoter sequences were compiled ready for the next stage of the workflow.

3.3.5 CpG island searching

The promoter sequences were then interrogated for the presence of CpG islands. A CpG island is defined as a region of the genome with at least 200bp, a GC content of greater than 50% and observed/expected CpG ratio of greater than 0.6. The Java program CpG island explorer (CpGIE) (Wang, 2004) is a useful analysis tool to check these features and was used in the present analysis. This CpGIE tool allows any sequence of interest to be checked for the presence of one or more CpG islands. The program has three settings for island length, GC content and observed/expected CpG ratio and these were set at 200bp, 50% and 0.6 respectively. The software then highlights the CpG island graphically and provides summary information regarding the start and end sites (Figure 3.4). Promoter regions that contained CpG islands progressed to the next stage of the process.

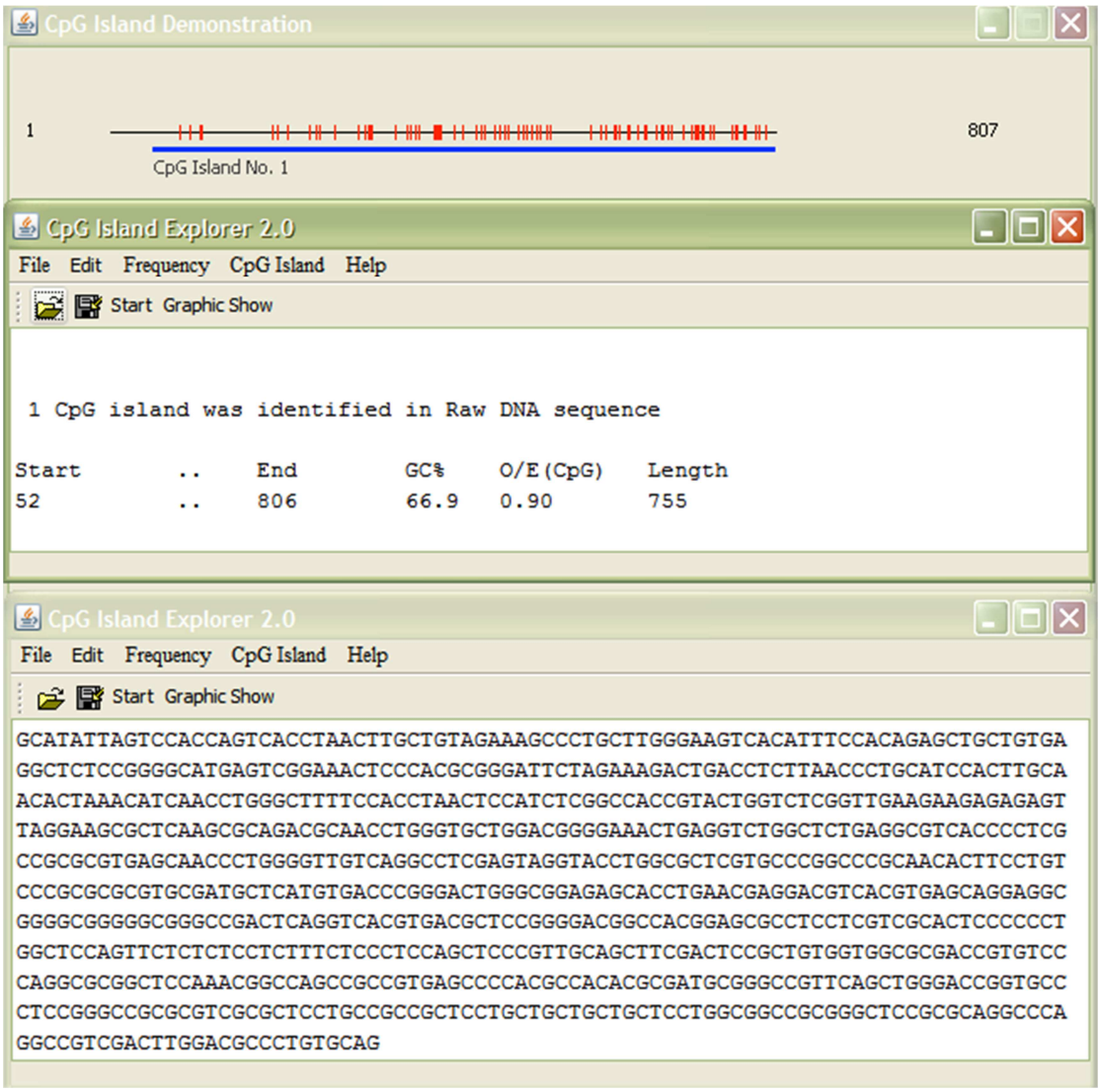


Figure 3.4 Screen capture from CpGIE application showing CpG Island search result.
The black line with red tags shows the sequence polled (*IGF2R* promoter) and positions of CpG sites. The blue line indicates the position of a CpG island within the promoter. The text shows the start and end of the CpG island within the polled sequence as well as the GC content and the observed/expected ratio.

3.3.6 *In silico* bisulphite modification

Bisulphite modification of DNA is a key step in many laboratory methods for assessing DNA methylation because it allows the pattern of methylation to be 'fixed' by effectively inducing a SNP at unmethylated cytosine residues. As a result the input material for most downstream methylation assays is bisulphite modified genomic DNA and it is to this modified sequence that primers for any subsequent PCRs are designed. The promoter sequences for selected genes were bisulphite modified *in silico* by highlighting all of the CG residues (CpG sites) in the sequence and replacing them with XG. This represents the presence of a methyl group on the cytosine residue blocking the bisulphite conversion.

Next all remaining C residues in the sequence were converted to T residues as this is what will occur when the bisulphite conversion reaction is carried out. Finally the XG placeholder was replaced with a C/TG, which represents the two possible alleles that may be present depending on the methylation status of the CpG site. Although this is a simple process in principle, in practice when working with large numbers of candidate genes and multiple promoters it can become labour intensive with the risk of introduction of errors at each step. To avoid this difficulty, a VBA macro for Word was written that performs all of the above steps at once as well as removing any hard line breaks from the sequence that can interfere with the assay design process.

3.3.7 Pyrosequencing assay design

The bisulphite modified promoters were analysed using the Qiagen PSQ Assay design software. The complete sequence for each gene of interest, including bisulphite-modification induced SNPs, was imported into the software package. All of the CpG sites within the sequence were assessed manually for assay suitability based on nucleotide spacing. Potential regions were selected based on a total amplicon length of 100bp or less and the presence of at least three CpG sites within the amplicons. Potential amplicons were then subjected to the assay design process (Figure 3.5). The assay design software created forward, reverse and sequencing primers based on optimum annealing temperatures and primer lengths, and minimisation of non specific binding (Figure 3.6).

Primer sets as generated by the assay design software were then sent for synthesis by Metabion AG (Martinsried, Germany).

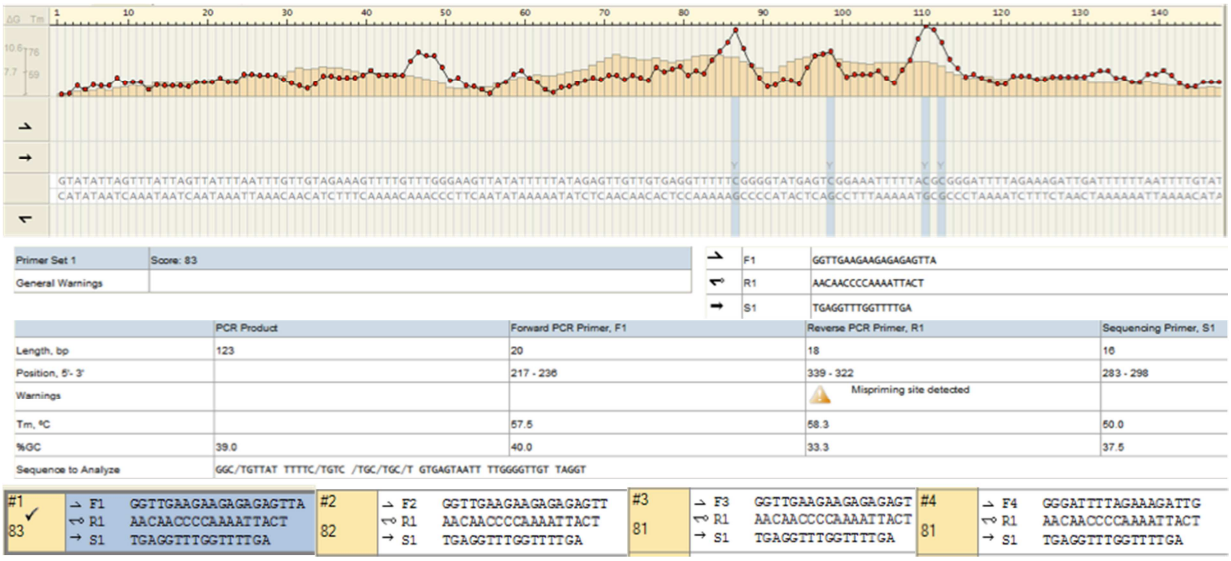


Figure 3.5 Screen capture from QCpG assay design application showing assay design result.
The red trace shows the GC content of the sequence shown. The sequence being assessed runs 5' to 3' left to right. CpG sites are indicated by blue bars. The forward, reverse and annealing primer sequences are show in the main box along with lengths and melting temperatures. The possible primer sets are shown at the bottom ranked by suitability score, with 100 being the highest possible score and indicating a very robust assay design.

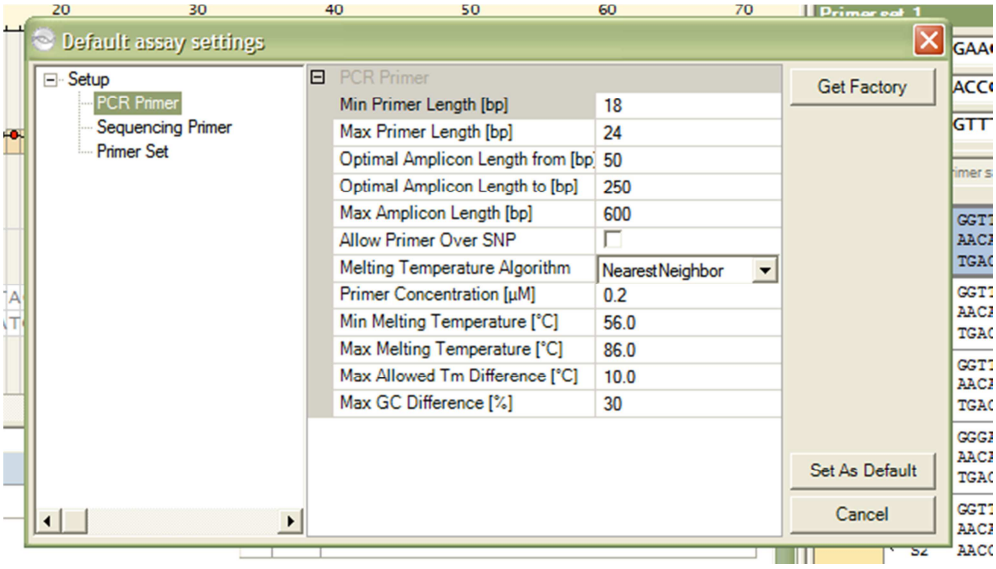


Figure 3.6 Screen capture from QCpG assay design application showing default assay settings.
The settings shown were utilised for all of the assay design runs and are the optimum conditions as chosen by the QCpG software.

3.4 Results

The raw expression dataset was analysed using the workflow shown in Figure 3.2. The results of each step of the analysis are detailed below.

3.4.1 Significance level - p value

The raw expression dataset comprised of a total of 19100 loci. This gene list was sorted by significance level with loci showing a significance level of greater than 0.05 excluded from the analysis. This resulted in the loss of 15400 genes at this stage with a total of 3700 genes moving to the next stage of the workflow (Figure 3.7).

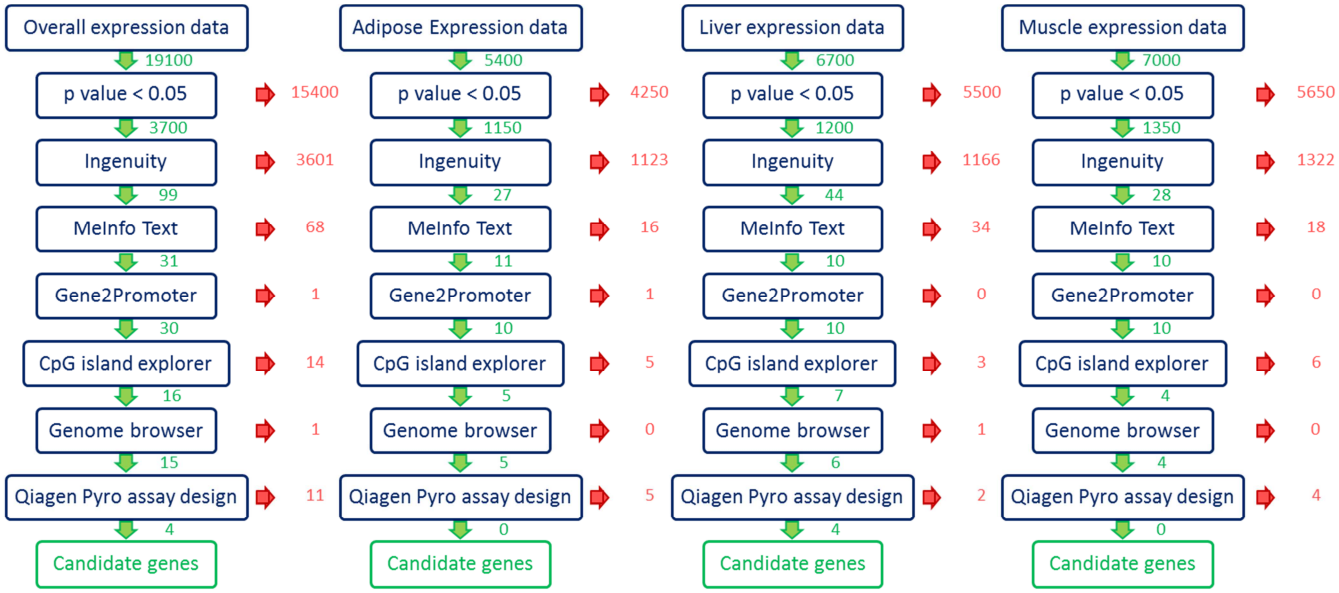


Figure 3.7 Candidate gene selection workflow showing attrition rate at each stage of the process, stratified for tissue type.
Genes lost at each stage are shown in red, genes progressing to the next stage of the analysis are shown in green.

3.4.2 Pathway analysis

The 3700 genes remaining in the analysis were assessed using Ingenuity Pathway Analysis as previously described. A total of 99 of these genes were shown to map significantly to the 'Knowledge base' across all six of the query pathways, with an attrition of 3601 genes at this stage of the workflow (Figure 3.7). Seventeen genes were mapped to the obesity pathway, 31 to cardiovascular disease, 13 to leptin signalling, 19 to insulin signalling and six to type 2 diabetes (Figure 3.8). A total of 13 of these 99 genes were mapped to the DNA methylation query pathway. The 99 genes that were shown to map to any of the above pathways progressed to the next stage in the workflow. These genes are summarised in Table 3.2.

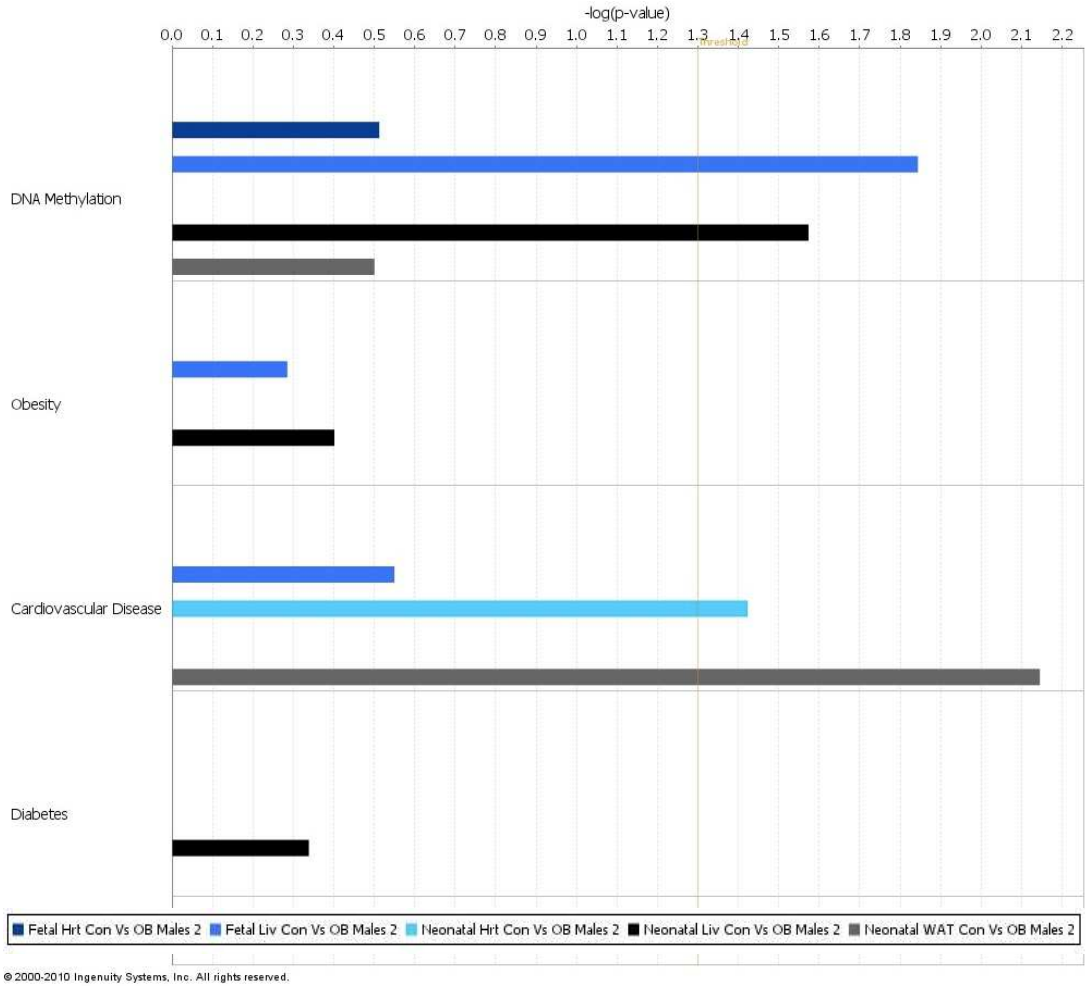


Figure 3.8 Screen capture from IPA application showing the mapping of candidates to obesity related pathways.
Coloured bars show the measure of probability that the polled gene set is related to the query pathways. Dark blue indicates foetal heart, medium blue indicates foetal liver, light blue indicates neonatal heart, black indicates neonatal liver and grey indicates neonatal white adipose tissue.

Outcome			Foetal Heart		Neonatal Heart		Foetal Liver		Neonatal Liver		Neonatal VAT	
1°	2°	Symbol	Fold Δ	p-value	Fold Δ	p-value	Fold Δ	p-value	Fold Δ	p-value	Fold Δ	p-value
CVD		APOA1			-8.70	0.04						
CVD		CD44									-1.29	0.03
CVD		CRP			-5.98	0.05						
CVD		DAG1									1.36	0.04
CVD		GJA1			1.36	0.04						
CVD		HMOX1			-2.02	0.05						
CVD		IL1B					2.65	0.01				
CVD		MEIS1									2.30	0.02
CVD		TLR2									1.63	0.01
CVD		AGT			-1.43	0.03						
CVD		CASQ2			1.54	0.04						
CVD		GAS6									1.79	0.02
CVD		PBX1			1.28	0.05					1.35	0.04
CVD		TAC1			1.22	0.04						
CVD		APOM			-10.98	0.04						
CVD		BMPR2									-1.05	0.04
CVD		F2RL3					1.32	0.04				
CVD		F3									2.90	0.02
CVD		LCP2					1.38	0.04	1.44	0.01		
CVD		MAP3K5									1.60	0.03
CVD		MAPK11					1.28	0.03				
CVD		MYL2									-119.17	0.00
CVD		NR1H3					1.68	0.02				
CVD		PTGIR			1.50	0.00						
CVD		RAG1					1.08	0.05				
CVD		RCAN1			1.82	0.01						
CVD		SERPING									1.48	0.02
CVD		TNFRSF1									4.41	0.04
Diab		FXN			1.19	0.05			1.31	0.04		
Diab		ICA1							1.27	0.04		
Diab		IFNGR1							1.59	0.01		
DNA-M		CTCF					1.38	0.04				
DNA-M		IGF2R							1.42	0.03		
DNA-M		PRMT7					1.34	0.02				
DNA-M		RB1							2.28	0.04		
DNA-M		RBL1							2.17	0.04		
DNA-M		RBL2	-1.12	0.03								
DNA-M		ATF7IP							1.62	0.02		
DNA-M		GATAD2A					1.24	0.04				
DNA-M		PICK1	-1.31	0.02								
DNA-M		PIWIL4					-1.13	0.02			-1.14	0.04
IGF1 Sig	DNA-M	FOS									2.18	0.04
IGF1 Sig		IGFBP4									2.32	0.01
IGF1 Sig		NRAS			-1.09	0.02						
IGF1 Sig		YWHAB							1.63	0.04		
IGF1 Sig		CSNK2A1			1.13	0.02					1.21	0.01
IGF1 Sig		CSNK2A2	-1.22	0.04								
IGF1 Sig		IGFBP5									2.08	0.04
IGF1 Sig		MITF			1.30	0.04						
IGF1 Sig		MRAS									1.36	0.02
IGF1 Sig		NEDD4									1.62	0.04
IGF1 Sig		PXN							1.33	0.03		
IGF1 Sig		RRAS2									1.80	0.01
Lep sig		AKT2					1.29	0.04				
Lep sig	IGF1 Sig	MAPK1	-1.08	0.01								
Lep sig	IGF1 Sig	AKT3	-1.12	0.05								
Lep sig		PLCD1	-1.25	0.02							1.56	0.04
Lep sig	IGF1 Sig	PRKAR1A			1.27	0.04					1.47	0.04
Lep sig		ADCY4			1.44	0.00						
Lep sig	IGF1 Sig	MAP2K2	-1.28	0.02								
Lep sig	IGF1 Sig	PIK3R2							1.25	0.03		
Lep sig		PLCE1			1.12	0.03						
Lep sig		PLCG2	-1.28	0.04			1.36	0.03				
Lep sig		PRKACA			1.23	0.04						
Lep sig	IGF1 Sig	PRKACB	-1.13	0.05								
Lep sig	IGF1 Sig	PRKAG1			1.18	0.04			1.33	0.03		
Lep sig	IGF1 Sig	PRKAG2									2.32	0.04
Obs		ADRB2					1.39	0.01				
Obs		ARNT							1.23	0.00		
Obs		CEBPB					1.96	0.03				
Obs		ESR1					-1.34	0.02			2.14	0.03
Obs	IGF1 Sig	MAPK8							1.11	0.02		
Obs		MPO							5.10	0.04		
Obs	Lep sig	SOCS3							1.54	0.05		
Obs	Diab	UCP2							2.37	0.04		
Obs		WNT10B							-1.25	0.02		
Obs		CLOCK							1.27	0.02		
Obs		FOXS1					1.46	0.02				
Obs		MTHFSD							1.30	0.01		
Obs		PIP4K2B					1.33	0.04			1.30	0.04
Obs		SDC3	-1.29	0.02			1.50	0.01				

Table 3.2 Candidate genes generated by IPA analysis.

Queries run along the far left. CVD – Cardiovascular disease, Diab – Type 2 Diabetes, DNA-M – DNA Methylation, IGF sig – Insulin signalling, Lep sig – Leptin signalling, Obs – Obesity. The tissues and timepoints run along the top of the table, showing fold change and p value for each candidate. Expression levels are colour coded with green indicating an upregulation and red indicating a downregulation. Some of the candidates are common to more than one pathway (for example MAPK1 is found in both the Leptin and Insulin signalling pathways).

3.4.3 Evidence of differential methylation

The 99 genes remaining in the workflow at this stage were assessed using the text data mining tool MIT as previously described. A total of 31 genes showed evidence of differential methylation leading to an attrition of 68 genes at this stage of the workflow. (Figure 3.7). These 31 genes progressed to the next stage of the workflow.

3.4.4 Promoter availability

The 31 genes remaining were then assessed for the presence of an experimentally verified promoter region using the Genomatix Gene2Promoter tool. In all but one of the genes analysed promoter sequences were located and verified (Table 3.7). This resulted in a total of 30 genes moving to the next stage of the workflow.

3.4.5 CpG island searching

The promoter regions generated by Gene2Promoter for all 30 of the remaining genes were assessed for the presence of CpG islands using the CpGIE tool previously described. Sixteen of the genes assessed showed at least one CpG island in their promoter, thus a total of 14 genes were dropped from the analysis at this stage (Figure 3.7). The 16 remaining loci moved to the next stage of the analysis.

3.4.6 SNP searching

The promoter regions of the 16 genes left in the analysis were assessed for the presence of single nucleotide polymorphisms (SNPs) that could interfere with primer binding and assay design in the next workflow step. One gene was found to have an extremely heterogeneous promoter region and so was excluded from the analysis at this stage (Figure 3.7). A total of 15 genes progressed to the final stage of the workflow.

3.4.7 Pyrosequencing assay design

The remaining 15 genes were assessed using Qiagen Pyrosequencing assay design software. Of the 15 genes remaining 11 showed a high CpG density within the promoter region that precludes their usefulness in pyrosequencing assay design. As such a total of four candidate genes were produced by the workflow (Figure 3.7). These genes were *Esr1*, *Fxn, Igf2r*, and

Rbl2. The promoters for each of the four genes in which assays were designed are shown in Figure 3.9 including full annotations. Both the *Esr1* and *Rbl2* genes were shown to have possible single nucleotide polymorphisms (SNPs) at their primer binding sites (Figure 3.9).

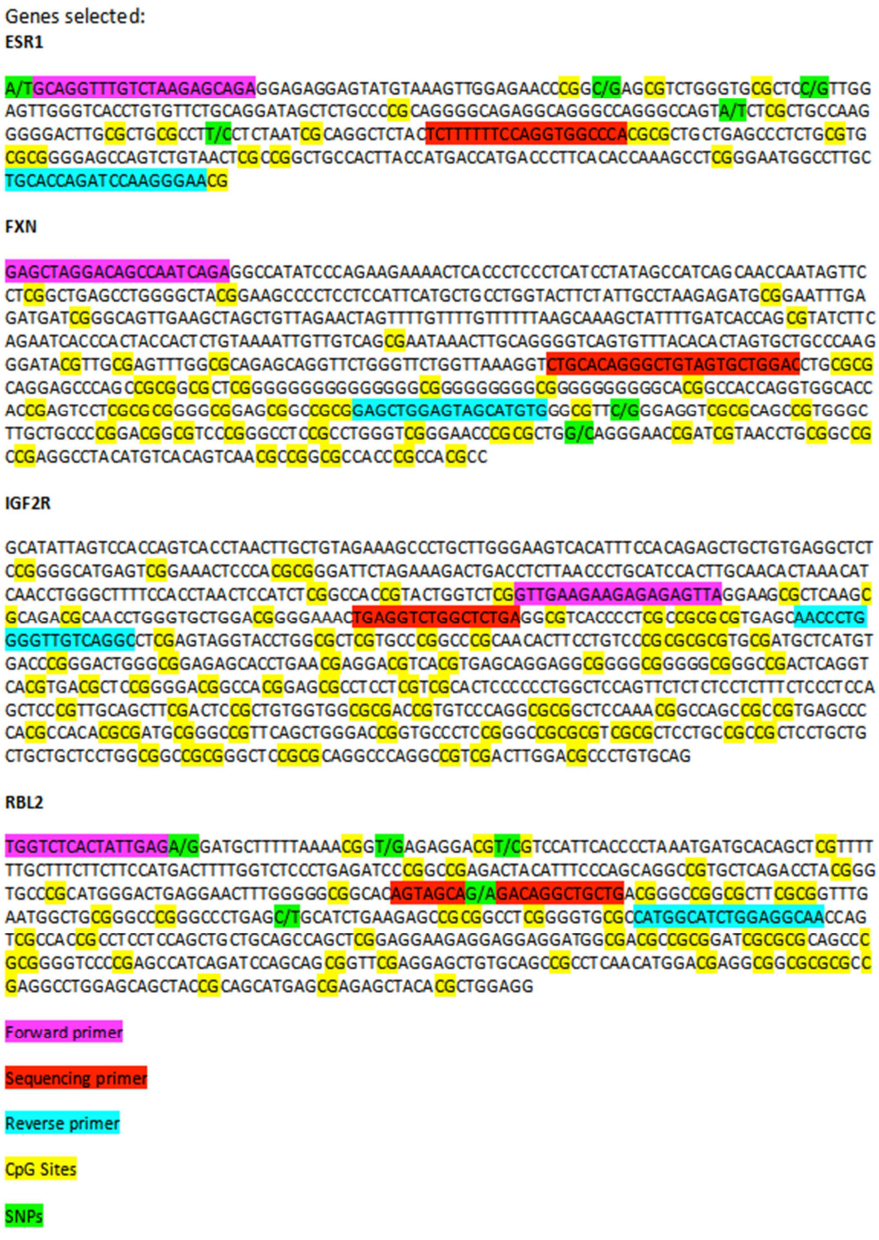


Figure 3.9 Promoter sequences for the four candidate genes including pyrosequencing assay designs as produced by the QCpG package.
The forward, sequencing and reverse primer annealing sites are shown highlighted in purple, red and blue respectively as indicated in the key above. CpG sites throughout the promoter are shown highlighted in yellow. Known SNP sites are shown in green indicating the major and minor alleles.

3.5 Discussion

The work described in this chapter aimed to identify a set of candidate genes from an input gene expression dataset that could subsequently be analysed for promoter methylation. This was achieved and resulted in the selection of four suitable candidates that were then assessed using Pyroseqeuncing as detailed in Chapter 4. The workflow presented in this chapter was developed iteratively and represents a novel method for the identification of candidate genes from large datasets. It has wide ranging applications as it can be utilised in the interrogation not only of gene expression datasets but indeed any gene list based dataset where the objective is to assess any possible effect of differential promoter methylation. One of the advantages of this workflow is that it is possible to adapt at various stages to change the filtering stringency. For example at the most basic level it would be possible to employ a more or less stringent p value cut off at the outset to allow more or less genes into the pathway analysis stage. Further, if regions of the genome other than Promoter regions or indeed CpG islands were of interest it would be possible to substitute these search tools for ones more appropriate to the research question.

As is the case with any such parsing methodology this workflow has inherent strengths and limitations which should be considered when evaluating how suitable this approach is for the identification of differentially methylated targets. The major advantage of this data processing method is that potential candidates are not dropped simply because they show a relatively small fold change in expression between the conditions of interest. This is because the pathway analysis approach used in this study included the entire list of genes that showed significant ($p<0.05$) differential expression between the two types of maternal nutrition i.e. standard chow and obesogenic diet. Utilising a traditional 2-fold cut off the total number of genes available for pathway analysis would have been 340 compared to the 3704 generated without a fold change cut off. The result is that fewer genes are excluded from the analysis in the early stages and therefore there is greater potential for the discovery of novel candidates that would otherwise have been discounted. Indeed, the greater the number of genes considered, the stronger the evidence that a pathway of interest is altered by our dietary intervention. If such a pathway consists of a large number of genes but only a small number make it through to the pathway analysis stage then the statistical ability to discern the effect of the intervention on the pathway is limited.

A further strength of this approach lies in the software used to perform the pathway analysis, or more specifically the 'knowledge base' which forms the backbone of the

software (Gusev et al., 2007). This manually curated database of gene and protein interactions draws its information from a large variety of sources including published scientific literature as well as KEGG (Kyoto encyclopaedia of genes and genomes) and a number of GO (gene ontology) databases. The broad spectrum of sources polled by this database and the regularity with which it is updated (weekly) ensures that pathway analysis queries are complete and include the most recently published findings relating to possible candidate genes. This could however also be seen as a limitation - pathway analysis protocols are limited by nature as they use prior knowledge to infer the likely molecular basis behind any changes observed. There is a possibility that novel findings are less likely as pathway programs poll prior published knowledge and if a novel gene or protein has not previously been shown to exert an effect in a context of interest it will not appear in the database.

Other limitations of this methodology include the attrition rate as the workflow progressed – at each stage of the analysis genes were lost so that at the final stage a relatively small number of candidates remained. Although this demonstrates the suitability of this protocol for the distillation of a large gene list into a set of targeted candidates it also illustrates the fact that a sufficient number of genes are required at the outset if targets are to be identified. That is to say that input gene lists in the order of 1000s of genes are required, encompassing most gene expression array datasets. This limits the usefulness of this workflow to large datasets. Indeed when this work flow was attempted on a smaller scale with an initial input of some 50 potential candidates selected with a fold change cut off of 2.0, the level of significance attainable in the IPA models were so low that it was not possible to discern which pathways were the most significant. Again this illustrates that an input gene list in the order or 1000s of genes is required for interrogation for this methodology to be useful. A major limitation of this methodology is in the number of genes that exhibit differential expression in response to the dietary intervention based on the P-value cut off of 0.05, but as previously mentioned this cut off could be revised if it proved unsuitable for an alternative application.

A final aspect of this method that is subject to some limitations is the Pyrosequencing assay design phase. During this process the promoter sequence for each potential candidate, having been identified as a CpG island and *in silico* bisulphite modified, is assessed for suitability for a Pyrosequencing PCR assay. Pyrosequencing PCRs (pPCRs) share common traits with more conventional PCR reactions in that the forward and reverse primers need to have similar annealing temperatures for the assay to be viable (Shen et al., 2007a). These

primers must be located within the sequence to be analysed at loci that do not span CpG sites, as overlapping a CpG site that has been bisulphite modified will likely result in poor interaction between the primer and the sequence (Colella et al., 2003). In the case of pPCRs this is further complicated by the presence of the sequencing primer that must also be located at loci that are not subject to CpG methylation. This combination of factors means that the likely success of assay designs are biased to CpG density within the promoter and therefore extremely CpG dense promoters may only yield one viable assay design, rather than multiple options. This problem is exacerbated with shorter promoters where the number of potential primer annealing loci is further reduced. The implications of this are that a limited number of CpG sites can be assessed within promoters. This problem could potentially be overcome through the use of alternative sequencing technologies such as bisulphite patch PCR (Varley and Mitra, 2010) and bisulphite sequencing (Sato et al., 2011), which allow greater coverage of CpG sites within a given promoter.

The approach used in this chapter is based on the premise that gene promoters are the likely site of functionally relevant differential methylation. There is however an increasing body of evidence suggesting that conserved gene sequences several kilobases up or downstream of promoter CpG islands, termed CpG island shores, harbour differential DNA methylation (Bell et al., 2011, Irizarry et al., 2009). There is also some evidence that these DNA methylation changes within CGI shores has functional consequences on gene expression levels (Cosgrove et al., 2011). As knowledge of the methylation variable regions of the genome increases it can inform the proposed workflow by shifting the focus from solely being on promoter regions to those flanking the promoter site and more distally.

Although IPA represents a powerful tool in the identification differentially methylated candidates from a gene expression array dataset there are numerous other packages which can be used to achieve a similar goal. One such package is DAVID (Database for Annotation, Visualization, and Integrated Discovery) (Da Wei Huang and Lempicki, 2008), a free web-based tool which was designed to explore interactions between molecules and to allow direct mapping of gene names to gene ontology data. Although a powerful tool in itself, DAVID is not a true pathway analysis package and also does not poll the same number of data sources as IPA. Other tools which can be used to investigate large gene expression datasets include Metacore (Ekins et al., 2006), PathVisio (van Iersel et al., 2008) and GenMapp (Salomonis et al., 2007). Other investigators have used custom pathway design packages such as WikiPathways to write their own pathways of interest before polling the more conventional databases (McKay et al., 2008).

In summary, the workflow presented in this chapter represents a novel methodology for the identification of differentially methylated target genes from an expression dataset. These genes were then taken forward for further analysis, in this case DNA methylation analysis by pyrosequencing as presented in Chapter 4.

Chapter 4: Is Differential DNA Methylation Programmed by Exposure to an Obesogenic Diet *In Utero*?

4.1 Background

The development of a candidate gene analysis pipeline as detailed in Chapter 3 resulted in the identification of four target genes to be analysed for gene specific DNA methylation. These genes were selected based upon their differential gene expression in liver tissue in the mouse model of diet induced obesity previously described (Samuelsson et al., 2008). The current chapter describes the quantification of DNA methylation at specific CpG sites in the four selected target genes; *Esr1, Fxn, Igf2r* and *Rbl2*.

Existing literature pertaining to the four target genes supports a potential role of epigenetic mechanisms in the regulation of their expression and in turn in the pathogenesis of obesity.

The first of the four candidate genes, *Esr1,* encodes for oestrogen receptor alpha (ERα), one of the two ER isoforms. ERα is a DNA-binding transcription factor that regulates gene expression, the principal ligand of which is 17β-oestradiol (oestrogen) (Gao et al., 2008). At low levels of circulating oestrogen ERα is predominantly confined to the cytosol, however as oestrogen is a steroidal hormone it is able to pass through cell membranes and bind to a receptor (Bjornstrom and Sjoberg, 2005). Oestrogen binding results in the migration of the receptor to the nucleus, receptor dimerization and binding of the receptor dimer to hormone response sequences of the DNA. The complex between DNA and ER then recruits the other transcription machinery and as a result proteins are expressed (Hall et al., 2001).

In addition to the identification of this locus in an animal model of over-nutrition, there is also a body of evidence from human studies that underscores the relevance of this gene to studies of the developmental programming of obesity. ER is most commonly associated with breast cancer, where in approximately 70% of cases it is found to be overexpressed (Ali and Coombes, 2000). This overexpression may result in the development of tumours by stimulating over proliferation of mammary cells resulting in accumulated DNA damage and disruption of the normal apoptotic processes. SNPs in *ESR1* have also been strongly implicated in the development of breast cancer (Stacey et al., 2007). More recently however

SNPs in *ESR1* have also been associated with the development of obesity (Chen et al., 2009a).

In the context of epigenetic changes *ESR1* is a key locus as it exhibits differential DNA methylation in disease cases compared to unaffected controls. Indeed DNA methylation changes in *ESR1* have been linked to a number of cancers including breast, (Widschwendter et al., 2004), prostate (Li et al., 2004) and colorectal cancer (Issa et al., 1994) as well as being linked with inflammation in ulcerative colitis (Tominaga et al., 2005). As such a decrease in expression levels of *ESR1* may program an obese phenotype due to aberrant oestrogen signalling. As such this combination of both a known link to obesity and differential DNA methylation makes *ESR1* a strong candidate for assessment.

The second candidate is *Fxn* (Frataxin), the human homologue of which (*FXN*) codes for a ~17 kDa protein the deficiency of which causes Friedreich Ataxia (FRDA), a neurodegenerative disorder that causes various systemic problems such as ataxia, loss of proprioception and cardiomyopathy. There are a number of common polymorphisms in *FXN* which give rise to this condition. Frataxin has been shown to have a variety of regulatory functions within the cell, such as iron homeostasis (Ramirez et al., 2010), regulation of respiration and control of antioxidant defences (Moreno-Cermeno et al., 2010). It has also been characterized as an iron chaperone (Cook, 2010). Defects in the *FXN* control mechanism result in metabolic disturbances caused by the build-up of iron in the mitochondria and increased oxidative stress (Patel and Isaya, 2001, Marobbio et al., 2011). Reduced expression of *Fxn* has also been recently associated with diet induced obesity in mice (Pomplun et al., 2007). Recent human studies have also shown that *FXN* exhibits differential DNA methylation in intron one which results in reduced gene expression and the development of FRDA (Castaldo et al., 2008), and as such *Fxn* represents a potentially interesting target in investigating the effects of DNA methylation on the pathogenesis of obesity.

The third candidate gene identified is *Rbl2,* a member of the retinoblastoma (RB) family of tumour suppressor genes. It binds members of the DNA binding E2F transcription factor family, regulating preadipocyte proliferation and differentiation. Adipocyte number is a major determinant of fat mass in human adults and polymorphisms in *RrlBL2* have been linked with aberrant control of adipocyte differentiation (Benetti et al., 2008).

DNA methylation has previously been described within the promoter of *Rbl2* (Al-Mahdawi et al., 2008), however these methylation changes have only been linked to Friedreich's

ataxia (Baghi, 2009). *Rbl2* is a member of the retinoblastoma (RB) family of tumour suppressors that binds members of the E2F transcription factor family, regulating gene transcription. It plays a key role in the regulation of pre-adipocyte proliferation and differentiation (Dimas et al., 2009). The ability of adipocytes to proliferate is key in the maintenance of fat mass in adults (Spalding et al., 2008). Thus increased expression levels of *Rbl2* may result in a decrease in adipocyte proliferation during development. Challenging these adipose cells with food high in fat and glucose may lead to cellular hypertrophy (Kubota et al., 1999), resulting in insulin resistance and increased fatty acid levels. This increase in circulating fatty acids might then exert effects both on the liver, resulting in non-alcoholic fatty liver disease and on peripheral tissues including the pancreas, causing type 2 diabetes (Kelley et al., 2003) (Boden, 2011). Although other epigenetic modifications have been described in the promoter of *Rbl2* (Wang et al., 2008) there is no current evidence that differential DNA methylation of *Rbl2* affects gene expression. The link between *Rbl2* and type 2 diabetes, a known sequela of obesity, makes *Rbl2* a potentially interesting target for downstream methylation analysis.

The fourth and final candidate gene selected from the expression array dataset is *Igf2r,* a maternally imprinted gene that codes for the Insulin like growth factor 2 receptor. This receptor forms part of the insulin-like growth factor system along with insulin-like growth factors 1 and 2 (*Igf1* and *Igf2*), the type 1 cell-surface receptor (Igf1r), the insulin receptor (*Ir*) and the circulating IGF-binding proteins (IGFBPs) (Jones and Clemmons, 1995). The actions of the IGFs are controlled by *Igf1r* and the insulin receptor and are involved in the moderation of cell growth and differentiation. *Igf2* is subject to further control by *Igf2r*, which as a cell surface receptor is able to traffic *Igf2* across the membrane, internalising it and making it available for degradation. *Igf2r* can therefore be classified as an inhibitor of cellular growth, with loss of function of *Igf2r* having been shown to promote foetal overgrowth (Lau et al., 1994) in mice. As an inhibitor of cellular growth *IGF2R* is a tumour suppressor gene and mutations in *IGF2R* have been associated with several human cancers including head and neck (Jamieson et al., 2003), lung (Kong et al., 2000) and breast (Oates et al., 1998). Changes in DNA methylation levels in the differentially methylated region 2 (DMR2) of *Igf2r* have been associated with decreased gene expression and result in foetal overgrowth in sheep (Young et al., 2001). As such decreased gene expression levels at the *Igf2r* locus may program an obese phenotype by disrupting the insulin signalling pathway. *Igf2r* therefore represents a potentially promising candidate for the assessment of DNA methylation in relation to the development of obesity.

4.2 Aims and objectives

The aim of this investigation was to test the hypothesis that maternal overnutrition during pregnancy alters DNA methylation in specific target genes in the offspring. Evidence of altered DNA methylation would then prompt further analysis to assess whether these changes mediate the development of an adverse cardiometabolic phenotype. To achieve this, methylation was quantified in DNA extracted from the livers of juvenile rodents. DNA methylation was quantified in several CpG sites in four genes prioritised through *in silico* appraisal of gene expression data. DNA methylation was analysed in relation to *in utero* exposure to a maternal obesogenic diet.

4.3 Methods

4.3.1 The Mouse Model

The animal model utilised for this study is detailed in Chapter 3 (Samuelsson et al., 2008) Female C57BL/6J mice were randomised onto either a standard mouse chow (7% simple sugars, 3% fat, 50% polysaccharide, 15% protein, energy 3.5 kcal/g, n=20), or a highly palatable obesogenic diet (10% simple sugars, 20% animal lard 28% polysaccharide, 23% protein, energy 4.5 kcal/g, n=30) from weaning. After six weeks on the diet, animals were mated and stayed on the obesogenic diet throughout gestation. Tissues were harvested from offspring of these animals at multiple time points postnatally. For the present study DNA was extracted from liver tissue of 8-week-old male offspring using the DNeasy blood & tissue kit (Qiagen, Crawley, UK). Offspring from six control animals and six high fat fed animals were assessed.

4.3.2 Bisulphite Modification

1μg DNA sample in 20ul of water was bisulphite modified using the EZ DNA Methylation Gold Kit (Zymo, UK) to convert unmethylated cytosine to uracil. 130μl of CT conversion reagent was added to the sample and run on a SENSOQUEST^{TM} labcycler using the following conditions: 98°C for 10 minutes, 64°C for 2.5 hours and held at 4°C. The samples were loaded into columns and were spun with 600μl of M-Binding Buffer. Samples were then washed with 100μl of M-Wash Buffer, desulphonated with 200μl of Desulphonation Buffer, washed twice with 200μl of M-Wash Buffer and eluted in 10μl of M-Elution Buffer.

4.3.3 Pyrosequencing PCR

1µg of genomic DNA was bisulphite modified using the EZ Methylation Gold™ Kit (Zymo, Cambridge, UK) according to the manufacturers' protocol. 1µl of bisulphite modified DNA was then amplified in a PCR reaction containing 12.5µl Hotstar TAQ Mastermix (Qiagen, UK), 2.5µl of 25mM $MgCl_2$ (Qiagen, UK), 3.0µl dH_2O, 0.5µl of 100pmol/µl forward primer and 0.5µl of 100pmol/µl reverse primer (one of which is biotin labelled). The Mastermix excluding $MgCl_2$ was made as follows: 12.5µl Hotstar TAQ Mastermix (Qiagen, UK), 5.0µl dH_2O, 0.5µl forward primer and 0.5µl reverse primer (one of which is biotin labelled). In each Mastermix, 2.5ng DNA was added to each well. Primers are detailed in Table 4.1.

Assay	Forward Primer	Reverse Primer	Sequencing Primer
Esr1	AGTAGGTTTGTTTAAGAGTAGA	TTCCCTTAAATCTAATACA	TTTTTTTTTTAGGTGGTTTA
Fxn	TTGTATAGGGTTGTAGTG	CACATACTACTCCAACTC	GGGTTGTAGTGTTGGAT
Igf2r	GGTTGAAGAAGAGAGAGTTA	AACAACCCCAAAATTACT	TGAGGT TTGGTTTTG
Rbl2	TGGTTTTATTATTGAGAGAT	TTACCTCCAAATACCATA	AGTAGTAGGATAGGTTGTTG

Table 4.1 Forward, reverse and sequencing primers for each assay.
All forward and reverse primers were diluted to a standard concentration of 100pmol/µl. The second and third columns show the primer sequences for the initial pyro PCR. The final column shows the sequencing primer added to the PCR mix for sequencing.

The PCR reactions were then run on a LabCycler (Sensoquest, UK) under the following reaction conditions:

95°C for 15 minutes

50 cycles: 95°C for 15 seconds

60°C for 30 seconds

72°C for 15 seconds

72°C for five minutes

4°C ∞

The pyrosequencing PCR samples were then cleaned up to single stranded DNA using the Vacuum Prep Workstation (VPW) (Biotage, UK). 10µl of PCR product was added to each well of a PCR plate. To this was added 2µl of Streptavidin Sepharose beads and 38µl of binding buffer. Volumes were then made up to 80µl with dH_2O. The plate was mixed vigorously for five minutes. Following mixing each well of a pyrosequencing plate (Qiagen, UK) was filled with 0.5µl of sequencing primer at 10µM. (Table 4.2) and 11.5µl of annealing buffer. The vacuum block tool was used to remove the PCR product and bead mix (the biotin labelled primer was bound to the bead) from the PCR plate, before a rinse with ethanol, denaturing buffer and a wash step and deposited in the pyrosequencing plate. The pyrosequencing plate was then incubated at 80°C for two minutes. The samples were then run in duplicate on a Pyromark MD Pyrosequencer (Biotage, UK). Any duplicate pairs not within 5% of one another were repeated. The dispensation order for each assay is displayed in Table 4.2. An example of the program output from the Pyrosequencer is shown in Figure 4.1.

Assay	Analysis sequence	Dispensation sequence
Esr1	C/TGC/TGTTGTTGAGTTTTTTGC/TGTGC/TGC/TGGGGAGTTAGTTTGTAATTC/TGTC/TGGTTGTTATTTATTATGATTATGAT	GTCTGTCGTGTGAGTTAGTCGTAGTCTGTCGGAGTAGTGTGATCAGTCGTG
Fxn	TTGC/TGC/TGTAGGAGTTTAGTC/TGC/TGGC/TGTTC/TGGGGGGGGGGGGGGG	GTAGTCTGTCGTAGAGTATGTCAGTCTGTCAGTC
Igf2r	GGC/TGTTATTTTTC/TGTC/TGC/TGC/TGTGAGTAATTTTGGGGTTGTTAGGT	AGTCGTGATTCTGTCAGTCTGTCGTG
Rbl2	AC/TGGGTC/TGGC/TGTTTC/TGC/TGGTTTGAATGGTTGC/TGGGTTC/TGGGTTTTGAG	GATCTGTCAGTCTGTTCAGTCGTGATGTAGTCTGTCGTT

Table 4.2 Dispensation orders for each assay.
The first column shows the sequence to be analysed by the pyrosequencer with C/T indicating the CpG sites within the assay. The second column shows the actual sequence of nucleotides dispensed, calculated algorithmically for optimum efficiency in each assay.

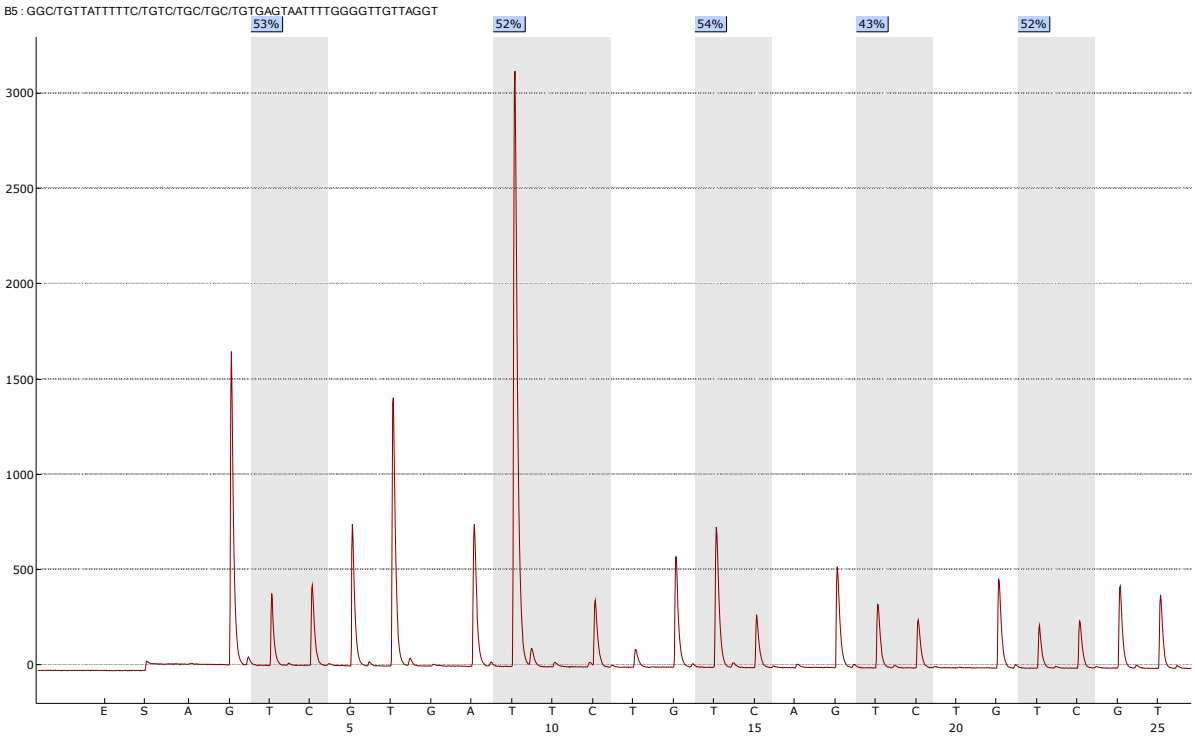


Figure 4.1 Pyrosequencer output for Igf2r assay.

The dispensation order as described in Table 4.2 is shown on the X-axis. Fluorescence is shown on the Y-axis. The red trace show the light emitted as each consecutive nucleotide is incorporated into the DNA strand. The grey highlighted areas show the CpG sites within the assay with the percentage methylation at each site in the light blue boxes. Site one is 53% methylated, Site two is 52% methylated, Site three is 54% methylated, Site four is 43% methylated and Site five is 52% methylated.

4.3.4 Validation

Pyrosequencing PCR reactions that were designed *de novo* required validation across a range of known DNA methylation concentrations to ensure that there was no bias in the amplification of methylated or unmethylated fractions of the PCR product. To achieve this 100% methylated and 0% methylated controls were produced. The 100% control was produced through use of the enzyme SssI Methylase to *in vitro* methylate mouse liver DNA samples. 4.4µl of DNA (for a DNA concentration of 2.5ng/µl), was combined with 2µl 1xBuffer 2, 6µl SssI Methylase, 6.6µl dH_2O and 1µl of 32mM SAM (diluted 1:8). Samples were then incubated at 30°C and SAM added every three hours for at least 16 hours. Respective samples were then pooled to make a stock. The 0% control was produced using whole genome amplification (WGA). Mouse DNA at a concentration of 50ng/µl was fragmented; a DNA library prepared, purified and amplified using the GenomePlex® Complete Whole Genome Amplification Kit (Sigma, Gillingham, UK). The DNA output was then purified using the DNA Clean and Concentrator-5 (Zymo, UK). Respective samples were then pooled to make a stock. These *in vitro* methylation (IVM) and whole genome amplification (WGA) controls were then mixed to produce a range of reference methylation percentages. To ensure that there was no preferential amplification of methylated or unmethylated DNA, reference mixes were produced both pre-PCR and post-PCR. These reference mixes are shown in Table 4.3. The pre and post PCR mixes were analysed by pyrosequencing in duplicate.

	Pre-PCR reference mix		Post PCR reference mix	
Methylation (%)	Volume IVM DNA (µl)	Volume WGA DNA (µl)	Volume IVM DNA (µl)	Volume WGA DNA (µl)
95	34.90 of 100%	1.90	87.40 of 100%	4.60
90	26.80 of 95%	1.50	67.00 of 95%	3.75
75	18.30 of 90%	3.70	45.75 of 90%	9.25
50	12.00 of 75%	6.00	30.00 of 75%	15.00
25	8.00 of 50%	8.00	20.00 of 50%	20.00
10	6.00 of 25%	9.00	15.00 of 25%	22.50
5	5 00 of 10%	5.00	12.50 of 10%	12.50

Table 4.3 Control DNA mixtures for pre and post PCR reference samples.
The bold values show the percentage methylation required. IVM – *In vitro* methylated DNA *WGA* – Whole genome amplified DNA. The volumes show the amounts of each mix required to produce the reference serial dilution.

4.3.5 Statistical Analysis

Quantification of DNA methylation by pyrosequencing generated an estimate of methylation at each specific CpG site expressed as a percentage measure for each DNA sample. All samples were run in duplicate and mean percentage methylation values were calculated. From this point onwards the mean of the duplicates will be referred to as the methylation level and mean methylation will represent the mean methylation across all CpG sites in any one amplicon. Given the expected non-normality of the methylation data (Figure 4.2) non-parametric analyses, specifically two-sample Wilcoxon rank-sum tests, were performed to compare mean percentages between experimental groups. Analyses were performed for each individual CpG site as well as the mean methylation level for each gene. Correlation between CpG sites in each gene was assessed using the Spearman's rank correlation test.

4.4 Results

4.4.1 Distribution of DNA methylation data

The use of non-parametric tests may have reduced the power to detect true associations if the distribution of methylation across these genes was in fact parametric. In the current dataset normality of the distributions was assessed by both visual inspection (Figure 4.2) and formally by the Shapiro-Wilk W test. However, the limited sample size makes interpretation of this latter test difficult. For instance, from inspection of the histograms, the methylation distributions appear skewed in three out of the four loci. However formal test of normality indicate no significant deviation from normality in any of the methylation measures. Evidence from data generated by myself and others in our group suggest that methylation data is generally not normally distributed (see subsequent chapters) and hence the decision was made to analysis the data using non-parametric tests (Figure 4.2).

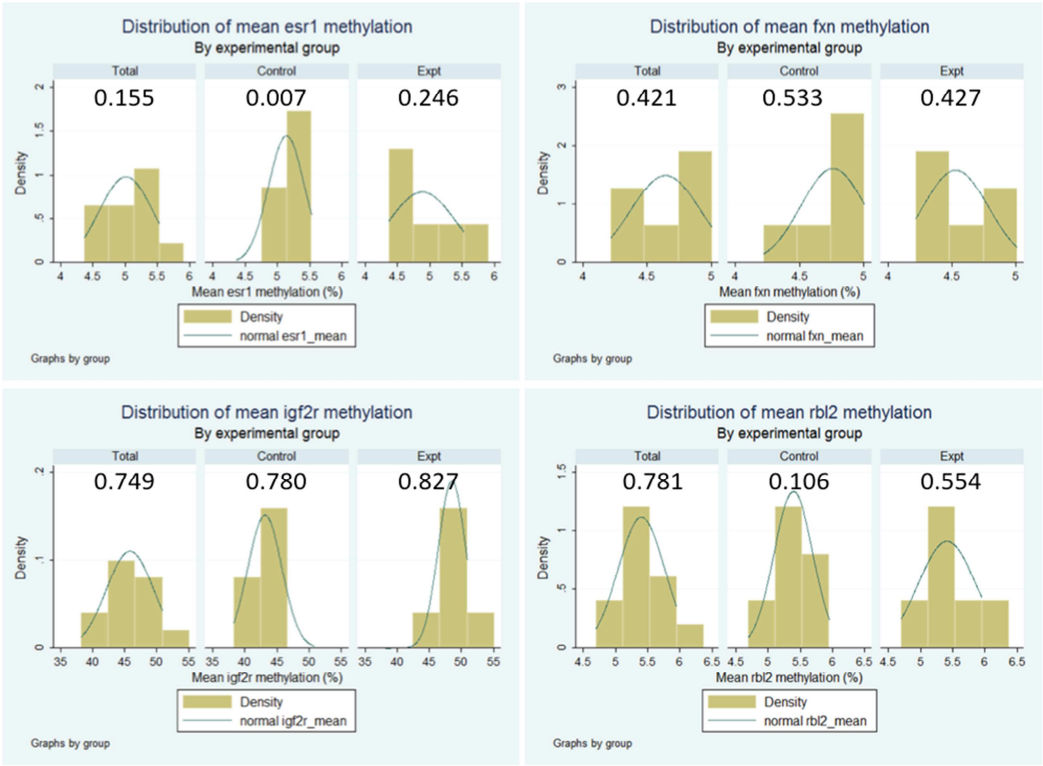


Figure 4.2 Distribution of mean methylation in each gene stratified by experimental group. The distribution of methylation data in each gene was assessed using the Shapiro-Wilk W test, all of the genes tested deviated significantly from a normal distribution other than *Igf2r.* Shapiro-Wilk p-values are quoted above each histogram. n = 12 for each separate histogram.

4.4.2 Validation of pyrosequencing assays

Validation was performed on all four of the target loci shown in Figure 4.2. The validation results for these assays are shown in Table 4.4. The validation was successful as the duplicates for each mixture were within 1.34 standard deviations, which is equivalent to 5% reproducibility. The validation results for both pre and post mix PCR were plotted as scatter graphs and a trend line and R^2 value were calculated (Figure 4.3). Both R^2 values were very close to one indicating that the assay is not exhibiting bias in amplification dependent on the methylation status of the template DNA. This provided confirmation that the assay provided accurate quantification of methylation in the case of all four assays so they were taken forward and performed on all 12 experimental samples.

Expected Methylation (%)	Observed Mean Methylation (%) in *Igf2r* assay		Observed Mean Methylation (%) in *Esr1* assay		Observed Mean Methylation (%) in *Fxn* assay		Observed Mean Methylation (%) in *Rbl2* assay	
	Pre-PCR	Post-PCR	Pre-PCR	Post-PCR	Pre-PCR	Post-PCR	Pre-PCR	Post-PCR
0	6.03	6.03	5.49	5.49	4.62	4.62	5.23	5.23
5	8.61	8.33	7.62	6.12	4.23	5.14	4.42	5.03
10	12.37	11.97	12.30	10.92	7.46	10.34	13.54	11.02
25	25.64	24.06	24.34	25.21	22.07	24.59	29.87	25.28
50	57.92	53.90	53.42	51.46	43.67	49.89	46.23	51.21
75	71.29	72.79	72.25	76.02	69.89	73.97	71.47	75.47
90	86.55	87.05	91.23	90.03	86.76	90.76	83.3	89.52
95	93.97	94.22	92.34	94.78	91.91	95.02	91.12	94.52
100	96.02	96.02	97.34	97.34	98.89	98.89	95.59	95.59

Table 4.4 Validation results for all four assays.

The expected methylation levels given the proportion of methylated to unmethylated DNA added are shown on the far left. The observed mean methylation values in both the pre and post-PCR reactions are shown in the subsequent columns.

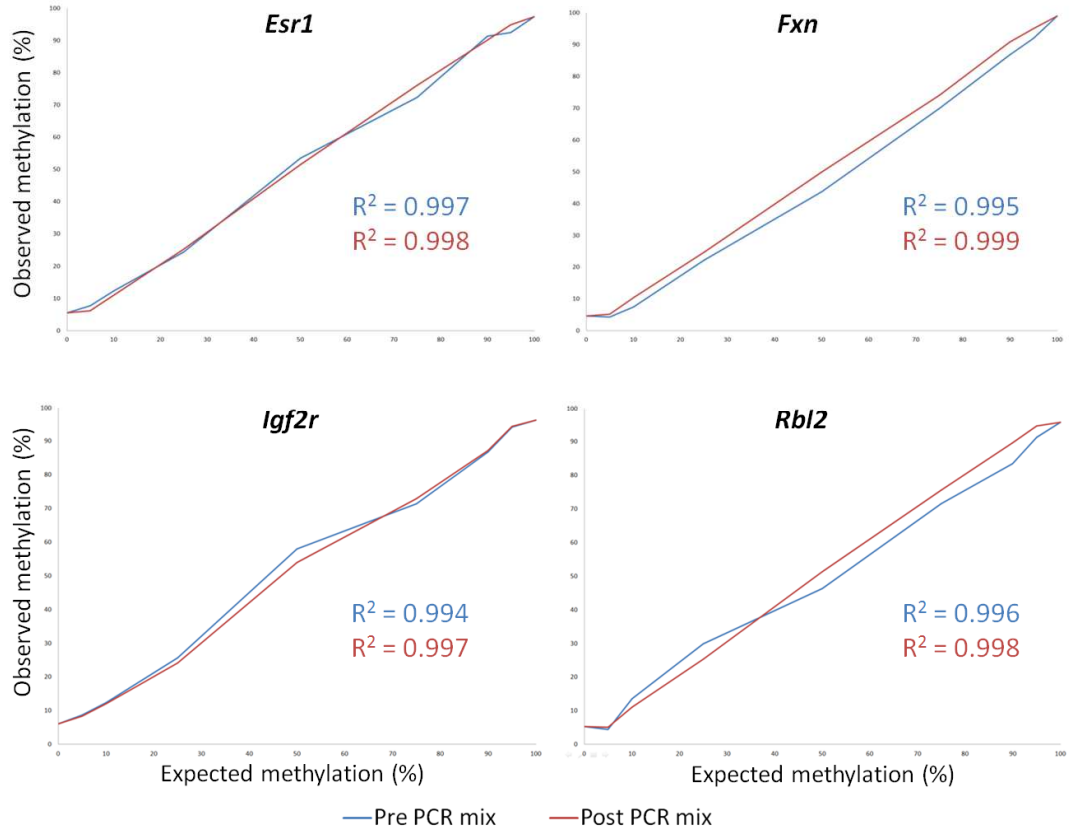


Figure 4.3 Validation curves for all four assays showing R^2.

The figures from Table 4.4 (above) are shown graphically here. The blue lines represent the pre-PCR mixes for each assay, the red lines the post-PCR mixes. The R^2 values as shown on the graphs indicate the degree of concordance between expected and observed methylation. The closer the R^2 value is to 1.0, the more closely they are correlated.

4.4.3 A comparison of gene specific methylation in experimental groups

A total of 25 CpG sites spanning the four candidate genes were successfully assessed by pyrosequencing.

Correlation between these CpG sites was assessed using Spearman's rank correlation. All five CpG sites assessed in *Igf2r* demonstrated strong positive pairwise correlations (Figure 4.4). In contrast, although some CpG sites spanning the *Esr1, Fxn* and *Rbl2* genes showed strong pairwise correlations, there was no consistent pattern demonstrated across all CpG sites measured within these genes (Figure 4.4). Interestingly, however, as shown in Figure 4.4 the correlation structure of the CpG sites spanning these three latter genes differed somewhat between control and exposed groups.

Esr1 showed a median (IQR) level of methylation across all six sites of 5.30% (4.80, 5.55) in control animals and 4.74% (4.49, 5.45) in exposed animals (Table 4.5). *Fxn* showed a median (IQR) level of methylation across all six sites of 4.79% (4.64, 4.99) (Table 4.5). *Rbl2* showed a median (IQR) level of methylation across all six sites of 5.38% (5.33, 5.54). *Igf2r* showed a median (IQR) methylation level of 44.04% (42.19, 44.92) (Table 4.5).

Table 4.5 presents median and IQR for methylation percentages measured at the individual CpG sites as well as the mean level across each amplicon. Overall, the median (IQR) methylation percentage across *Esr1, Fxn, Igf2r* and *Rbl2* was 5.08% (4.69, 5.33), 4.69% (4.40, 4.85), 45.47% (44.04, 48.54) and 5.42% (5.26, 5.65), respectively. With the exception of those mapping to *Igf2r*, no individual CpG site nor mean measure demonstrated significant differences in methylation levels between experimental groups (Table 4.5). For *Igf2r*, animals exposed to an obesogenic environment showed consistently higher methylation levels at each CpG site as well as the overall mean value (Table 4.5 and Figure 4.5). These differences were shown to be statistically significant (Wilcoxon test for mean methylation: $z = -2.88$, $p = 0.004$).

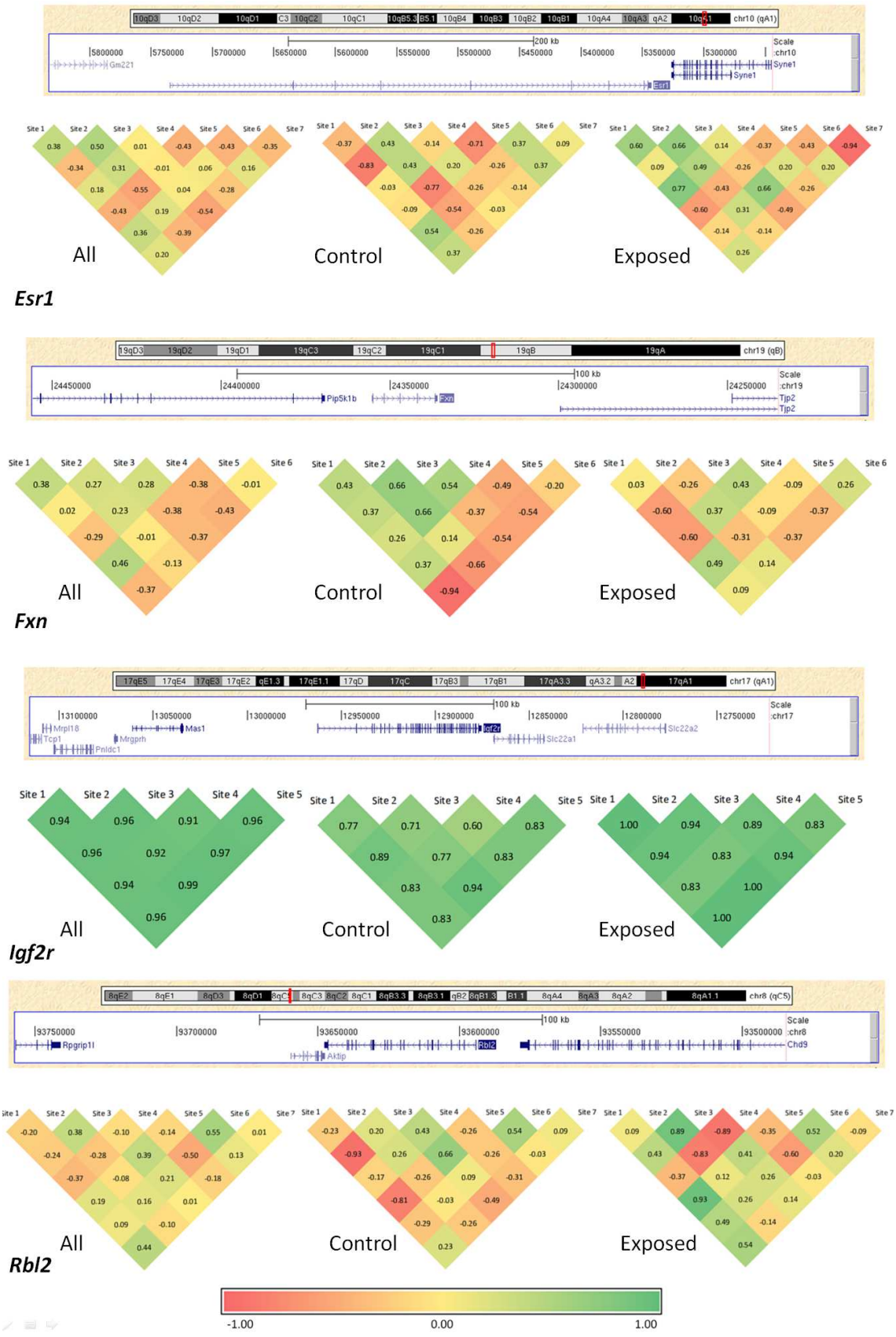


Figure 4.4 Correlation between methylation at each CpG site stratified by exposure group. Spearman's rank correlation coefficient values (rho) shown. 1.00 is highly positively correlated -1.00 is highly negatively correlated, colour ranked accordingly.

	All animals		Control		Exposed		Group comparison	
Gene (Site)	n	Median (IQR) Methylation (%)	n	Median (IQR) Methylation (%)	n	Median (IQR) Methylation (%)	Wilcoxon rank-sum z	p-value
Esr1								
Site 1	12	4.67 (3.36, 5.69)	6	5.40 (3.50, 5.55)	6	3.86 (3.23, 5.83)	0.16	0.873
Site 2	12	4.67 (3.50, 5.66)	6	4.67 (3.71, 5.74)	6	4.42 (3.44, 5.57)	0.64	0.522
Site 3	12	4.46 (3.40, 6.62)	6	5.80 (3.55, 7.15)	6	3.95 (3.26, 5.17)	0.96	0.337
Site 4	12	4.53 (3.94, 5.53)	6	5.07 (3.82, 6.37)	6	4.53 (4.15, 4.80)	0.48	0.631
Site 5	12	5.03 (4.22, 6.16)	6	5.58 (4.57, 6.07)	6	4.26 (3.66, 6.26)	0.80	0.423
Site 6	12	5.49 (4.52, 5.97)	6	5.41 (4.84, 5.78)	6	5.64 (4.20, 6.91)	-0.45	0.631
Site 7	12	6.13 (4.98, 6.80)	6	5.68 (3.71, 6.18)	6	6.80 (5.19, 6.97)	-1.60	0.423
Mean	12	5.08 (4.69, 5.33)	6	5.30 (4.80, 5.33)	6	4.74 (4.49, 5.45)	0.64	0.522
Fxn								
Site 1	12	7.07 (6.54, 7.51)	6	7.43 (6.81, 7.56)	6	6.63 (6.14, 7.35)	1.60	0.109
Site 2	12	6.76 (6.01, 7.22)	6	7.16 (6.10, 7.33)	6	6.72 (5.60, 6.76)	1.44	0.150
Site 3	12	3.39 (2.83, 4.37)	6	3.97 (3.07, 4.52)	6	2.96 (2.67, 3.91)	1.28	0.200
Site 4	12	3.36 (2.57, 4.11)	6	2.89 (2.54, 3.55)	6	3.98 (2.81, 4.23)	-1.28	0.200
Site 5	12	3.62 (2.82, 4.41)	6	3.62 (2.71, 4.49)	6	3.62 (3.37, 4.06)	0.00	1.000
Site 6	12	4.26 (2.90, 4.49)	6	4.41 (2.89, 4.50)	6	3.57 (2.91, 4.44)	0.48	0.631
Mean	12	4.69 (4.40, 4.85)	6	4.79 (4.64, 4.99)	6	4.47 (4.36, 4.81)	1.28	0.200
Rbl2								
Site 1	12	5.00 (4.11, 6.31)	6	5.32 (4.35, 6.32)	6	4.97 (3.93, 5.66)	0.64	0.521
Site 2	12	4.93 (3.90, 6.86)	6	4.40 (3.40, 5.19)	6	6.04 (4.67, 6.94)	-1.44	0.150
Site 3	12	3.39 (3.23, 4.29)	6	3.72 (3.22, 4.72)	6	3.39 (3.28, 3.84)	0.32	0.749
Site 4	12	5.96 (4.95, 7.96)	6	6.41 (5.13, 7.64)	6	5.63 (3.47, 8.28)	0.48	0.631
Site 5	12	4.98 (4.48, 6.56)	6	4.81 (4.73, 6.39)	6	5.83 (4.24, 6.56)	-0.32	0.748
Site 6	12	5.95 (5.84, 6.22)	6	5.91 (5.83, 5.98)	6	6.03 (5.85, 6.31)	-0.56	0.575
Site 7	12	6.18 (5.78, 6.78)	6	6.14 (5.83, 6.70)	6	6.40 (5.74, 6.85)	-0.32	0.749
Mean	12	5.42 (5.26, 5.65)	6	5.38 (5.33, 5.54)	6	5.43 (5.20, 5.76)	-0.32	0.749
Igf2r								
Site 1	12	48.46 (47.19, 51.34)	6	47.19 (45.02, 48.09)	6	51.34 (49.11, 53.04)	-2.88	0.003
Site 2	12	48.26 (45.98, 50.83)	6	45.98 (44.43, 47.46)	6	50.83 (49.32, 52.79)	-2.72	0.006
Site 3	12	46.69 (45.04, 50.31)	6	45.04 (43.93, 45.94)	6	50.31 (48.05, 52.20)	-2.72	0.006
Site 4	12	38.55 (36.03, 39.89)	6	36.03 (34.66, 37.01)	6	39.89 (39.04, 41.51)	-2.72	0.006
Site 5	12	46.47 (45.03, 50.33)	6	45.03 (42.84, 45.82)	6	50.33 (48.24, 52.72)	-2.88	0.003
Mean	12	45.47 (44.04, 48.54)	6	44.04 (42.19, 44.92)	6	48.54 (46.72, 50.80)	-2.88	0.003

Table 4.5 DNA methylation across all assays and all sites stratified by experimental group.
Median and IQR (Interquartile range) of percentage methylation shown for all of the samples combined and stratified by experimental group. Comparison between the exposure groups performed by Wilcoxon rank-sum test.

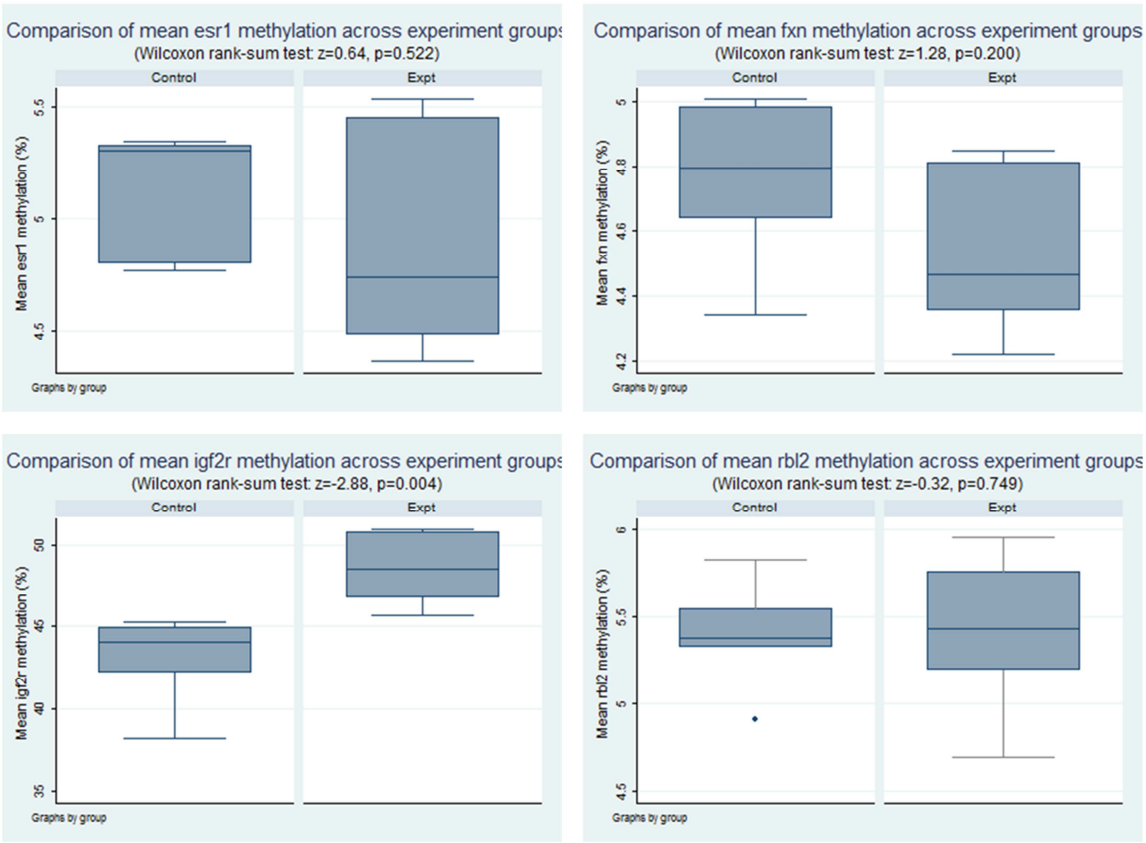


Figure 4.5 Mean methylation in DNA from livers of offspring from dams fed control and obesogenic diets before and during pregnancy.

Box and whisker plots shown for all four assays stratified by experimental group. Differences between experimental groups were assessed using Wilcoxon rank-sum test.

4.5 Discussion

Four candidate genes selected by *in silico* analysis of gene expression data (described in Chapter 3) were assessed for gene specific DNA methylation levels and levels compared between control and experimental groups. Three of the four candidate genes analysed showed generally low levels of DNA methylation with *Esr1*, *Fxn* and *Rbl2* showing median methylation levels of approximately 5% (range 2.54% to 7.56%). There was also no significant difference between DNA methylation at levels at these loci (as appraised by pyrosequencing) in control versus exposed animals. This does not preclude the role of DNA methylation in transcriptional regulation in these genes, rather it shows that in the case of the CpG sites assessed, DNA methylation is at a relatively low level and therefore one might infer that this is unlikely to be exerting an effect on gene expression. There may be higher levels of DNA methylation at other CpG sites within the promoters of these genes, or indeed other regulatory regions, which down regulate gene transcription and this could be assessed using alternative technologies which allow more extensive fine mapping of methylation patterns across the locus such as Sequenom's MassArray technology (Gloss et al., 2011).

Esr1 showed relatively low levels of methylation in its promoter region. It is interesting to note this given that in breast cancer models *Esr1* has been shown to be hypermethylated, albeit in tumour tissue (Ramos et al., 2010) and the analysis conducted in this study was undertaken in peripheral blood DNA. This suggests that *Esr1* has a role in the pathogenesis of cancer, a logical conclusion given that *Esr1* is a tumour suppressor gene (Suga et al., 2008). However *Esr1* has also been implicated in the pathogenesis of obesity due to the key role it plays in the oestrogen signalling pathway. Oestrogen signalling is involved in the control of food intake and silencing of *Esr1* in the hypothalamus (Musatov et al., 2007) and has been shown to lead to the development of the metabolic syndrome. The oestrogen signalling pathway, in which *Esr1* plays a pivotal role, has also been shown to exert an effect in liver tissue, as oestrogens influence glucose metabolism through the activation of *Esr1* (Riant et al., 2009). Hence, low levels of methylation of the *Esr1* promoter might be expected to lead to increased expression of *Esr1*. Further exploration of methylation across this 400kb locus is required to definitively assess the link between *Esr1* methylation and obesity.

The role of Frataxin (*Fxn*) is partially understood and it is known to be implicated in the binding and chaperoning of iron groups in the mitochondria, a role that is essential for

metabolism and the prevention of oxidative stress (Bulteau et al., 2004). Frataxin deficiency in the pancreas has however been shown to induce diabetes (Ristow et al., 2003) and reduced expression of Frataxin has also been shown to exacerbate obesity in a rodent model subjected to an obesogenic diet. (Kuhlow et al., 2010, Pomplun et al., 2007). Thus, decreased levels of DNA methylation in the *Fxn* promoter could plausibly be expected to have a functional influence via increased levels of *Fxn* expression. However in these data methylation levels observed in the *Fxn* amplicon analysed were too low to discern a meaningful relationship between DNA methylation at this locus, maternal nutrition and obesity.

In contrast to those genes discussed above, *Igf2r* showed median methylation levels of approximately 50% and also showed significant differences in DNA methylation between the two exposure groups (Table 4.5, Figure 4.5). This suggests that an increased level of DNA methylation is exerting an effect on phenotype via modulation of gene expression levels. The level of methylation observed in *Igf2r* is in line with current knowledge that the *Igf2r* locus is maternally imprinted (Sandovici et al., 2003, Latos et al., 2009). There is however potential for sex-specific effects of variable methylation at the *Igf2r* locus. Given that these data included only male animals it would be interesting to see if the same relationship is observed in female offspring. The key function of *Igf2r* is as an anti-proliferative agent, and it achieves this through clearing *Igf2* from the circulation (Braulke, 1999). *Igf2r* knockout mice tend to be much larger than wild type mice (Lau et al., 1994). Increased DNA methylation in the differentially methylated regions of *Igf2r* might lead to a decrease in expression of *Igf2r*, resulting in a decrease in circulating *Igf2* and therefore increased growth and the development of an obese phenotype.

The correlation structure of site-specific DNA methylation differed between control and exposed groups (Figure 4.6). This may be indicative of site-specific differences in DNA methylation between the groups. Alternatively, this may be a feature of random variation in DNA methylation between individuals that is unrelated to exposure status. Larger sample sizes would be helpful in gaining a better understanding of DNA methylation correlation structure. Understanding the short- and long-distance relationship between methylation at specific CpG sites is crucial to understanding the functional importance of this form of epigenetic variation (Bell et al., 2011). Where possible one should always consider the local methylation patterns rather than isolated CpG sites, this is reflected in the use of mean methylation levels across assays rather than locus specific results.

These data provide an interesting starting point in the investigation of the developmental programming of obesity. Demonstrating a causal link between gene specific differential DNA methylation and gene expression is a plausible target for future research. A rodent model with multiple time points and tissue types would be ideal, in which animals could have DNA and RNA extracted at birth and at various time points to assess the temporality of any changes in DNA methylation and if they map to gene expression changes and indeed phenotype. (Plagemann et al., 2009)

The data presented show that gene specific methylation measurements the *Igf2r* promoter differed between exposure groups in 8-week-old rodents. In attempting to identify the mechanisms underpinning the developmental programming of obesity (Remacle et al., 2011) it is clear that the temporality of any change in DNA methylation signatures is an important consideration. The current findings do not allow conclusions to be drawn regarding whether the observed changes in *Igf2r* methylation were present at birth or whether they arose secondary to other physiological changes consequent upon the maternal obesogenic environment. A key question is whether DNA methylation at this locus is affected by early life exposures such as the nutritional insult to which these animals were exposed (Sebert et al., 2011) and if so whether these methylation changes persist through early life and into adulthood. To investigate this DNA methylation analysis of the tissue of interest would have to be undertaken at multiple time points including in the foetus, neonate and later in life. This would allow the stability of epigenetic marks to be considered in light of the emergence of programmed phenotypic changes.

Given these findings it would be interesting to replicate this experiment in a human cohort study. As is often the case with translational studies however there is an issue of tissue comparability that should be considered. As both the gene expression and DNA methylation analysis in these rodents was performed on liver tissue, these data only can only inform tissue specific conclusions (Thompson et al., 2010a). In order for this experiment to be replicated in humans liver tissue would be required for DNA/RNA extraction. This could take the form of a liver biopsy but this invasive procedure may preclude the collection of liver samples, especially multiple samples collected over a period of time. An alternative would be to repeat this experiment in rodents using DNA extracted from peripheral blood to establish if methylation levels are similarly altered or at least reflective of the target tissue in a sub-set of loci. This would then allow peripheral blood samples to be used in humans to enable a translational study.

There are a number of points which need to be considered in relation to this dataset, including for example the fact that only male offspring were assessed in both the initial gene expression and the following DNA methylation analyses. The reasoning behind this is sound, with commonly observed sex discordance in programmed effects in animal models (McKay et al., 2011b, Khan et al., 2003, Samuelsson et al., 2008). However it would be interesting to perform a similar experiment in both male and female animals to observe sex specific differences in gene specific DNA methylation, particularly given the role of oestrogen in the mechanism of action of *Esr1* and the imprinted nature of the *Igf2r* locus.

Although the links between DNA methylation and gene expression have been well characterised (Deaton and Bird, 2011), DNA methylation is only one of many forms of regulation of gene expression, so lack of findings in this regard do not detract from the potential importance of differential gene expression or other epigenetic mechanisms which may be involved in regulating this.

Also, despite an attempt to target the *in silico* analysis to highlight differentially expressed genes which were also implicated as differentially methylated targets, a number of the genes selected showed overall low levels of methylation. Although this does not preclude these genes showing developmental programming effects, as more bioinformatic resources become available differentially methylated targets should become easier to locate. Currently there is no centralised repository of DNA methylation data that makes searching for information on the methylation status of a gene a labour intensive process. However with future advances in text data mining and information integration this methodology should become more commonplace.

In summary all four genes interrogated for promoter methylation using pyrosequencing showed some level of methylation. However only the *Igf2r* promoter, the locus in an imprinted gene, showed differences between the two experimental groups. This suggests that maternal overnutrition during pregnancy may affect DNA methylation in the offspring in a gene specific manner. This gene would benefit from more thorough investigation using higher resolution sequencing in a larger number of samples to determine if DNA methylation in *Igf2r* is robustly implicated in the programming of obesity.

Chapter 5: Is Differential DNA Methylation Associated with Adiposity in Childhood?

5.1 Background

Changes in epigenetic patterning and particularly changes in the levels of DNA methylation can have a profound effect on an individual's phenotype (Feinberg, 2007) because of the impact of these changes on gene expression. However, there is still limited empirical data, outside the cancer field, to support this widely held postulate. To gain a deeper understanding of the aetiology of pathologies such as developmentally programmed obesity one must first appreciate the mechanisms involved, including epigenetics. Numerous studies have identified, or suggested, factors which may alter patterns of DNA methylation in experimental models (Burdge et al., 2005, Waterland et al., 2007, McKay et al., 2011a, Mathers et al., 2010) These interventions include nutritional insults *in utero*, such as the depletion of maternal folate supply (McKay et al., 2011b), as well as maternal environmental exposures such as polycyclic aromatic hydrocarbons (PAHs) found in cigarette smoke (Herbstman et al., 2009). However, there is currently very little published literature which directly links changes in exposure to these environmental factors with altered epigenetic patterning and, in turn, with perturbed adiposity.

In considering developmental programming of disease the term 'critical window' is commonly used to describe periods in the developmental process during which key exposures can influence an individual's phenotypic outcome (Symonds et al., 2006). Gestation represents such a critical window because in very early life many physiological systems become 'hard-wired' and there is some evidence that appetite levels may be programmed during this life-stage (Bouret, 2009). This programming may arise through epigenetic marks being 'captured' and perpetuated, eliciting their effect on gene expression at a later stage in the life course (Mathers and McKay, 2009). Alternatively epigenetic markings established in early life may have a more proximal effect on gene expression, with the resultant altered gene expression inducing cumulative physiological effects during the life course. Life course approaches have been applied widely in epidemiology and represent a domain of the epidemiology field in their own right (Ben-Shlomo and Kuh, 2002, Kuh et al., 2003, Liu et al., 2010b). These approaches now routinely involve making measurements in a cross section of the population to assess the presence of an epigenetically induced change in physiological measurement of biomarker (Bjornsson et al., 2008). These measurements may be circulating concentrations of lipids or hormones or alternatively may

involve the assessment of samples of DNA or RNA. Of particular interest to us are measurements of epigenetic modifications, namely aberrant DNA methylation (Wang et al., 2010a). The question that can then be addressed is 'do early life events mark the epigenome with later consequences for childhood or adult health?' (Waterland, 2009). This can be achieved by analysing epigenetic patterns at a single time-point and relating these to later outcomes. Exposures preceding the time at which DNA was sampled can then be assessed to see if they influence epigenetic patterns. Such approaches rely on the assumption that the sample in question has been captured within a 'critical window' (Symonds et al., 2006). As epigenetic patterns change over time (as do many life course exposures such as infection and nutrition) a single sample snapshot may not provide the full picture (Christensen et al., 2009). However, it may provide some insight into the relationship between early life exposures, epigenetic patterns and adiposity (Lillycrop and Burdge, 2010). Much more work is required to establish not only the critical windows of exposure (current evidence pointing to *in utero* and early postnatal life) but also the critical windows of epigenetic plasticity (Vickers, 2011).

Epigenetic modifications provide a mechanism whereby evidence of the experience of early life exposures are 'captured' by the genome and exert effects on gene expression and health in later life. The use of animal studies including an early life nutritional insult is one way of assessing if obesity risk can be 'programmed' (Weaver et al., 2004). One such model exposes rodents to overnutrition *in utero* which results in an obese phenotype in later life (Samuelsson et al., 2008). In early life, these animals show altered levels of expression in genes implicated in the pathogenesis of obesity. It is plausible that these differential levels of expression may be regulated by epigenetic processes, namely DNA methylation (Movassagh et al., 2010). Many genes involved in relevant pathways such as appetite regulation and fat deposition are CpG dense and these so called CpG islands (CGIs) can harbour differential levels of methylation (Shen et al., 2007b). If a CGI is located in the promoter region of key regulatory gene then differential methylation may have functional consequences (Palou et al., 2011). Altered levels of DNA methylation in regulatory regions such as the transcription start site (TSS) could alter the tertiary structure of the complex making it more difficult for transcription factors and other transcription machinery to bind (Kass et al., 1997). Alterations in the binding of transcription machinery affect gene expression (Sengupta et al., 2003, Palacios et al., 2010).

Insulin-like growth factor 2 (*IGF2*) is part of the IGF gene family that encode a group of proteins expressed predominantly in liver but also at lower levels in other tissues. *IGF2*

plays a role in mammalian growth by influencing foetal cell division and differentiation and possibly metabolic regulation (Constancia et al., 2002). *IGF2* is an imprinted gene, that is a gene in which only one of the two parental alleles of a locus is expressed (Wang et al., 2010b). This mechanism is controlled by DNA methylation patterns in the *IGF2*-H19 region and targeted disruption of the *IGF2* gene in mice (Kalscheuer et al., 1993) shows that the paternally expressed *IGF2* gene is essential for normal embryonic growth. Aberrant expression of the *IGF2* protein has been linked with a number of conditions including Beckwith–Wiedemann syndrome (Eggenschwiler et al., 1997), a syndrome characterised by neonatal overgrowth and the risk of developing Wilms' kidney tumours. Loss of *IGF2* imprinting has also been observed in a number of cancers (Uribe-Lewis et al., 2011) (Woodson et al., 2004, Feinberg, 1999, Chen et al., 2000).

Recent gene-association studies have also linked polymorphisms in the *IGF2* gene to body weight (Gomes et al., 2005). The *IGF2* gene is located close to the insulin gene on chromosome 11p in humans and this region of the genome has been strongly implicated in the regulation of childhood and adult body weight and fat mass (Gaunt et al., 2001, Rodriguez et al., 2004, Ukkola et al., 2001, Zhang et al., 2010b).

In addition to evidence that *IGF2* methylation is associated with various phenotypic traits, there is also emerging evidence to suggest that early life exposures can influence methylation at this locus (McKay et al., 2011b, Gong et al., 2010, Zhang et al., 2010b). Of particular interest is the fact that malnutrition in early life may precipitate changes in *IGF2* methylation measured in later life. Recent work on the Dutch Hunger Winter cohort showed that individuals undernourished in the periconceptional period exhibit lower *IGF2* methylation levels in blood cells at age 60 years, when compared with their unexposed same-sex siblings (Heijmans et al., 2008). These findings make *IGF2* an ideal candidate for further analysis of the role of aberrant DNA methylation in the developmental programming of obesity.

The *TACSTD2* gene encodes for the Tumour-Associated Calcium Signal Transducer 2, a protein that transduces intracellular calcium signals and acts as a cell surface receptor (Tsujikawa and Tano, 2007). Mutations of this gene are most strongly associated with gelatinous drop-like corneal dystrophy (GDLD), an autosomal recessive disorder characterized by severe corneal amyloidosis (Zhang and Yao, 2010). However recent studies in our group showed that catch up growth is associated with both *TACSTD2* DNA methylation and gene expression at age 12 years. Catch-up growth occurs in young children after a period of growth deficit when the insult causing the deficit is removed. It consists of

a period of abnormally high growth followed by a progressive slowing of growth rate until the normality has been reached (Williams, 1981). There is evidence that catch-up growth is associated with the development of obesity in both human studies (Ong et al., 2000) and mice (Ozanne and Hales, 2004). Interestingly, the studies in our group also showed that differential DNA methylation in the *TACSTD2* locus was also associated with childhood phenotypic traits including weight, waist, HDL and total cholesterol and fat mass (Groom et al., 2012).

Thus, two candidate genes that have been implicated in childhood obesity were interrogated to assess whether DNA methylation in these genes was associated with adiposity in childhood.

5.2 Aims and objectives

The aims of this chapter were to assess it the gene loci *IGF2* and *TACSTD2* exhibited differential DNA methylation within their promoter regions. It also set out to test if any such differential DNA methylation was associated with phenotypic markers of metabolic health including weight, bioimpedance and BMI at age eight years.

5.3 Methods

5.3.1 The Gateshead Millennium Study

The Gateshead Millennium Study (GMS) (Parkinson et al., 2011) recruited 1029 mothers living in Gateshead, North East England who gave birth to children between June 1999 and May 2000. These women were invited to take part in two main studies: a Feeding and Growth Study and an Iron Deficiency Study. Initially parents completed a series of questionnaires relating to feeding behaviour, development and illnesses. Each child also attended a health check at 13 months. Extensive data on early growth were collected. Two further questionnaires were then undertaken, one at 30 months and another at five to six years. A further 619 of the children and their parents were revisited between October 2006 and December 2007 to obtain anthropometric and physical activity measurements, to complete a further questionnaire and to collect saliva samples for DNA extraction. A sub-set of these samples and information were utilised for the study described here. The sub-set did not differ from the complete cohort in any of the parameters analysed with the exception of age in the *IGF2* sub-group who were on average 0.17 years younger than the cohort as a whole (Table 5.1).

5.3.2 DNA extraction

Saliva samples were collected from both mothers and children using the Oragene DNA kit (DNA Genotek, Ontario, Canada). Samples were incubated at 50°C overnight before being separated into two 1ml aliquots for biobanking and three 500ul aliquots for DNA extraction. 20□l of Oragene DNA Purifier was added to each DNA aliquot before a 10-minute incubation on ice. Samples were then spun at 13,000rpm for five minutes, the supernatant transferred to a new 1.5ml tube before addition of 500□l of room temperature 100% ethanol. Samples were mixed gently by inversion and incubated at room temperature for 10 minutes. Samples were then centrifuged at 13,000 rpm for two minutes. The supernatant was discarded and the pellet washed with 250□l 70% ethanol. The pellet was then air dried before resuspension in 100□l Tris EDTA buffer. Following an overnight incubation at room temperature the samples were quantified on a NanoDrop 1000 Spectrophotometer (Thermo, UK).

5.3.3 Exposure and phenotypic data

Early life exposures that might plausibly influence DNA methylation patterns at age eight years were included in the current study. Gestational age (weeks), sex and birth weight (g) were recorded from delivery records. Birth weight z-score was derived by subtracting mean birth weight for the cohort from measured birthweight for the individual and dividing this result by the standard deviation of the cohort (Parkinson et al., 2011). Thrive index (TI) as a measure of conditional weight gain was derived from algorithms derived by earlier work on this study cohort conducted by Wright et al (Wright et al., 1998, Wright et al., 1994). The thrive index methodology uses the weight of the infant during the early weeks of its life as a baseline with which to compare subsequent weights which have been adjusted for regression to the mean. This requires the transformation of weight measures into standard deviation scores using a computer algorithm and allows the weight gain of infants who are not following usual growth trajectories to be more effectively monitored. Maternal age, height, weight and BMI and paternal age, height, weight and BMI were collected at the 2006/2007 follow-up clinic visit. Height was measured to 0.1 cm using a Leicester portable height measure with the head in the Frankfort plane. Weight was measured to 0.1 kg and bioimpedance was also measured using TBF-300MA scales (Tanita Corp., Japan). BMI was calculated from the measured height and weight for each individual.

5.3.4 Pyrosequencing of *IGF2* and TACSTD2 loci

1µg of genomic DNA was bisulphite modified using the EZ Methylation Gold™ Kit (Zymo, Cambridge, UK) according to the manufacturers' protocol. 1µl of bisulphite modified DNA was then amplified in a PCR reaction containing 12.5µl Hotstar TAQ Mastermix (Qiagen, UK), 2.5µl of 25mM $MgCl_2$ (Qiagen, UK), 3.0µl dH_2O, 0.5µl of 100pmol/µl forward primer and 0.5µl of 100pmol/µl reverse primer (one of which is biotin labelled). The Mastermix excluding $MgCl_2$ was made as follows: 12.5µl Hotstar TAQ Mastermix (Qiagen, UK), 5.0µl dH_2O, 0.5µl forward primer and 0.5µl reverse primer (one of which is biotin labelled). In each Mastermix, 2.5ng DNA was added to each well. Primers are detailed in Table 5.1 and their localisation in the context of the genes shown in Figure 5.1

Assay	Forward Primer	Reverse Primer	Sequencing Primer
IGF2	AGTAAGAAATTGGATAGG	AAACCCCAACAAAAACCACT	TTTTTTAGGAAGTATAGTTA
TACSTD2	CTAGGTACTGTACTGTCA	ACTCACTAGTACGACAATA	TCTAACCAGGTAATTGTCCAC

Table 5.1 Forward, reverse and sequencing primers for both assays.
Forward and reverse primers were diluted to a standard concentration of 100pmol/µl

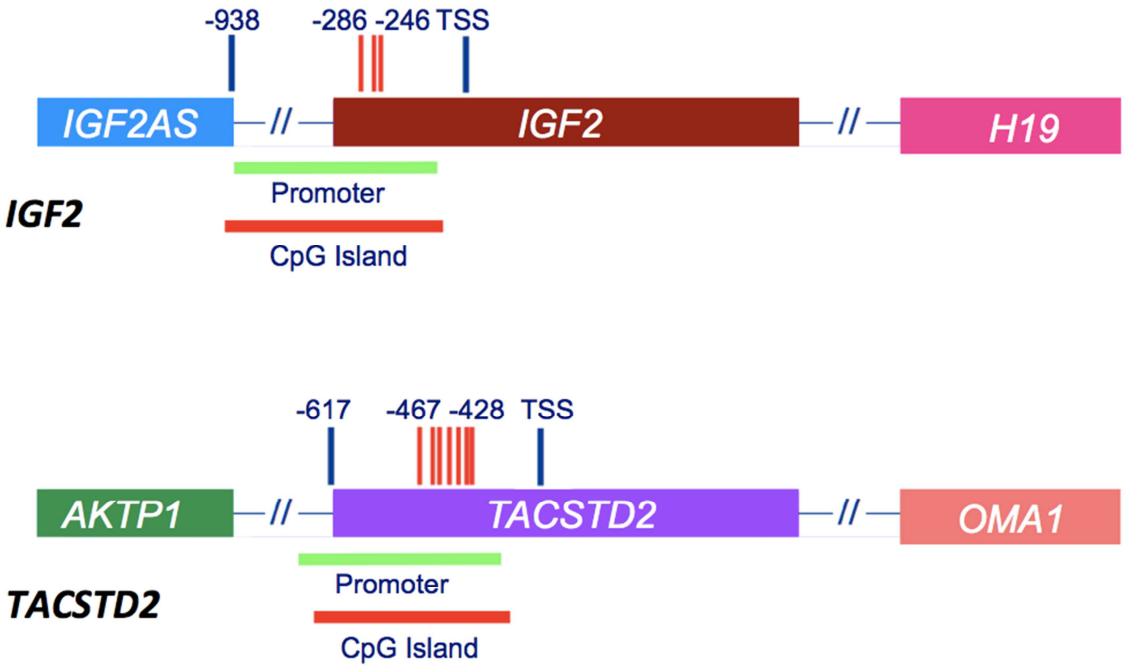


Figure 5.1 Positions of CpG sites within *IGF2* and *TACSTD2* amplicons.
Promoter is indicated in green, CpG island in red. The *TACSTD2* amplicon contains seven CpG sites indicated in red, the *IGF2* amplicon contains three CpG sites.

The PCR reactions were then run on a LabCycler under the following reaction conditions:

95°C for 15 minutes

50 cycles: 95°C for 15 seconds

60°C for 30 seconds

72°C for 15 seconds

72°C for five minutes

4°C ∞

The pyrosequencing PCR samples were then cleaned up to single stranded DNA using the Vacuum Prep Workstation (VPW) (Biotage, UK). 10µl of PCR product was added to each well of a PCR plate. To this was added 2µl of Streptavidin Sepharose beads and 38µl of binding buffer. Volumes were then made up to 80µl with dH_2O. The plate was mixed vigorously for five minutes. Following mixing each well of a pyrosequencing plate (Qiagen, UK) was filled with 0.5µl of sequencing primer at 10µM and 11.5µl of annealing buffer. The vacuum block tool was used to remove the PCR product and bead mix (biotin labelled primer bound to bead) from the PCR plate, before a rinse with ethanol, denaturing buffer and a wash step and deposited in the pyrosequencing plate. The pyrosequencing plate was then incubated at 80°C for two minutes. The samples were then run in duplicate on a Pyromark MD Pyrosequencer (Biotage, UK). Any duplicate pairs not within 5% of one another were repeated. An example of the program output from the Pyrosequencer is shown in Figure 5.2.

5.3.5 Statistical analyses

Pyrosequencing generated an estimate of methylation at each CpG site expressed as a percentage for each DNA sample. Samples were run in duplicate and mean percentage methylation values calculated. From this point the mean of the duplicates will be referred to as the methylation level and mean methylation will represent the mean methylation across all CpG sites in any one amplicon. Analyses were performed for each individual CpG site as well as the mean methylation level for each gene. Correlation between CpG sites in each gene was assessed using the Spearman's rank test. Data was tested for normality using the Shapiro-Wilk normal data test. The representative nature of the

subgroups was tested using the Wilcoxon rank-sum test. The association between DNA methylation and phenotypic traits were assessed by multiple regression adjusting for age and sex. Further analysis was performed by Spearman's correlation stratifying for sex.

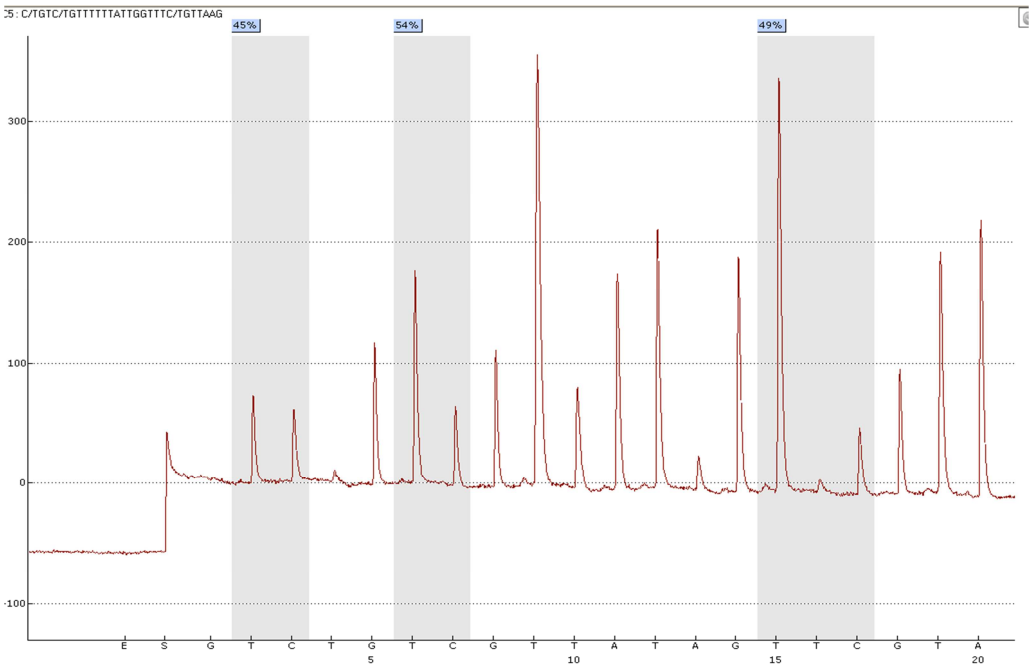


Figure 5.2 *IGF2* Pyrogram.
Note the sequence being analysed on the X-axis. The grey bars indicate the CpG sites within the gene, the red peaks the light emission as each consecutive nucleotide is incorporated into the DNA strand. In this sample Site one is 45% methylated, Site teo is 54% methylated and Site three is 49% methylated.

5.4 Results

In a cross-sectional study DNA samples collected from children in the Gateshead Millennium Study (GMS) were analysed to assess whether DNA methylation at age eight was associated with indices of body composition at the same age.

5.4.1 Cohort details and sample selection

Individuals were chosen at random from the GMS based on the availability of DNA samples. Gene specific DNA methylation analyses for *IGF2* and *TACSTD2* were performed separately at different points in time so the overlap between samples analysed in both was not complete. The overlap between samples assessed with respect to DNA methylation across the two loci is summarised in Table 5.2. A total of 132 children (68 male, 64 female) were analysed with respect to *IGF2* methylation while a total of 90 children (45 male, 45 female) were analysed with respect to *TACSTD2* methylation. To ensure that these sub-sets were representative of the whole cohort the outcome variables

were assessed using the Mann-Whitney U test or T test depending on normality of the data (Table 5.2). The majority of outcome variables in either the *IGF2* or *TACSTD2* sub-group did not differ significantly from the whole cohort with the exception of age in the *IGF2* sub-group who were on average 0.17 years younger than the cohort as a whole (Mann-Whitney U z = 3.42, p=6.00E-4). The *TACSTD2* sub-group was on average 2.33cm taller than the whole cohort (T test z=-3.90, p=0.0001). Thus the two sub-groups differed between one another with respect to age and height (Table 5.3).

With the exceptions of bioimpedance in the *IGF2* sub-group (Shapiro-Wilk t = -2.40, p=0.018) and bioimpedance in the *TACSTD2* sub-group (Shapiro-Wilk t = -2.00, p=0.048), there were no significant differences between the two subgroups for any of the measured phenotypic characteristics (Table 5.3).

	Sub-group A		Sub-group B		Comparison‡	
Phenotypes at age eight years‡	n	Median (IQR)	n	Median (IQR)	Stat (z or t)	p-value
IGF2* Sub-group vs. Remaining GMS Cohort						
Age (years)	132	7.33 (7.00, 7.58)	477	7.50 (7.17, 7.83)	3.42	6.00E-04
Height (cm)†	132	124.80 (120.40, 128.68)	467	125.25 (121.60, 129.15)	1.10†	0.272
Weight (kg)	131	25.30 (22.45, 28.70)	466	25.58 (22.75, 29.1)	0.41	0.684
Bioimpedance (Ω)	132	624.25 (587.75, 668.75)	462	632.00 (592.00, 676.00)	1.11	0.266
BMI	131	16.48 (15.52, 17.49)	466	16.29 (15.20, 17.92)	-0.36	0.719
TACSTD2* Sub-group vs. Remaining GMS Cohort*						
Age (years)	90	7.58 (7.25, 7.83)	519	7.42 (7.08, 7.75)	-1.43	0.152
Height (cm) †	90	127.13 (123.50, 131.00)	509	124.80 (121.00, 128.50)	-3.90†	0.0001
Weight (kg)	90	25.68 (21.50, 34.50)	507	25.50 (23.00, 28.40)	-1.43	0.153
Bioimpedance (Ω)	89	650.50 (581.50, 713.50)	505	628.50 592.00, 668.00)	-1.76	0.078
BMI	90	14.96 (14.31, 20.81)	507	16.31 (15.51, 17.5)	0.19	0.849
TACSTD2* Sub-group vs. *IGF2* Sub-group**						
Age (years)	61	7.67 (7.25, 7.92)	103	7.25 (7.00, 7.67)	-3.43	0.0006
Height (cm) †	61	127.90 (124.30, 131.00)	103	124.60 (119.40, 128.25)	-4.00†	0.0001
Weight (kg)	61	28.80 (21.80, 34.60)	102	25.40 (23.40, 28.35)	-1.71	0.088
Bioimpedance (Ω)	60	647.25 (579.25, 712.25)	103	623.50 (588.00, 657.00)	-1.77	0.078
BMI	61	18.34 (14.31, 21.26)	102	16.49 (15.73, 17.30)	-0.01	0.993

Table 5.2 Comparison between sub-groups in overall cohort.

‡ Median, lower and upper percentiles are presented and comparisons tested by Wilcoxon rank-sum, unless otherwise stated. † Mean and standard deviations are presented and comparisons tested by T-test. * 132 children from the GMS were utilized for analyses in *IGF2* methylation, hence comparisons were made between the remaining participants. ** 90 children from the GMS were utilized for analyses in *TACSTD2* methylation, hence comparisons were made between the remaining participants. ** 29 children were common to both gene-specific analysis sub-groups, hence comparisons were made between the remaining.

	Boys only		Girls only		Normalcy test‡	Sex comparison	
Phenotypes at age eight years	n	Median (IQR)	n	Median (IQR)	p-value	Statistic (z or t)	p-values
***IGF2* sub-group**							
Age (years)	68	7.38 (7.04, 7.58)	64	7.33 (7.00, 7.67)	0.454	0.14†	0.893†
Height (cm)	68	126.15 (119.15, 130.25)	64	124.58 120.83, 127.35)	0.367	1.07†	0.288†
Weight (kg)	67	25.30 (22.45, 29.4)	64	25.85 (22.58, 28.58)	0.001	-0.01*	0.995*
Bioimpedance (Ω)	68	617.25 (577, 647.75)	64	634.50 (597.75, 692.25)	0.056	-2.40†	0.018†
BMI	67	16.27 (15.52, 17.15)	64	16.74 (15.4, 17.74)	0.001	-0.95*	0.348*
***TACSTD2* sub-group**							
Age (years)	45	7.42 (7.17, 7.75)	45	7.58 (7.33, 7.83)	0.861	-1.35†	0.181†
Height (cm)	45	127.90 (122.90, 130.95)	45	126.15 (123.70, 131.00)	0.806	0.31†	0.754†
Weight (kg)	45	26.00 (21.50, 36.25)	45	25.00 (21.45, 34.00)	1.00E-5	0.38*	0.702*
Bioimpedance (Ω)	45	622.50 (556.00, 696.00)	44	679.25 (592.00, 741.25)	0.412	-2.00†	0.048†
BMI	45	14.97 (14.54, 20.51)	45	14.90 (14.08, 20.81)	1.00E-6	0.58*	0.564*

Table 5.3 Phenotypic variables measured in children at age eight years.
Data shown for all of the individuals studied and stratified by sex. ‡ Shapiro-Wilk data normal test, †T test, *Mann-Whitney U test.

5.4.2 *IGF2* methylation

Methylation at three CpG sites in *IGF2* was assessed using Pyrosequencing in children from the GMS cohort as described above. Correlation between CpG sites in *IGF2* was assessed using Spearman's rank correlation. *IGF2* showed strong positive correlation between CpG sites with rho values of greater than 0.60 (Figure 5.3). This indicated that mean methylation across all three sites could be used in downstream analysis without significant loss of sensitivity as mean methylation across all three sites was a representative measure of the separate methylation values. *IGF2* showed a median level of methylation across all three sites of 42.72 (38.72, 47.31) (Table 5.4).

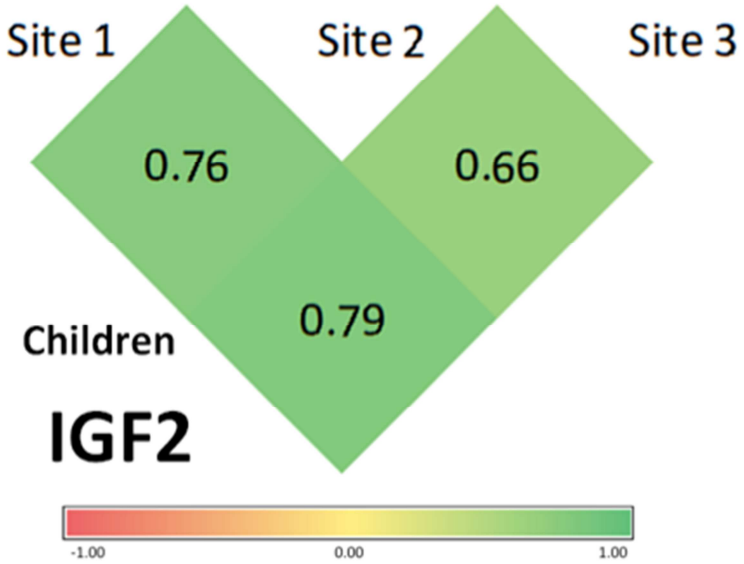


Figure 5.3 Chromosomal position of *IGF2* gene and correlation between CpG sites.
Degree of correlation is shown by colour coding, positive correlation is shown in green.

	Boys only		Girls only		Sex comparison	
***IGF2* CpG Site**	n	Median % Methylation (IQR)	n	Median % Methylation (IQR)	Statistic (z)	p-value
Site 1	68	38.55 (32.77, 43.86)	64	38.21 (33.92, 46.4)	-0.72	0.472
Site 2	68	43.66 (39.1, 48.32)	64	43.82 (40.865, 49.11)	-0.54	0.591
Site 3	68	44.74 (40.39, 49.71)	64	45.57 (40.94, 50.21)	-0.87	0.385
Mean	68	42.01 (38.98, 46.19)	64	43.33 (38.49, 47.73)	-0.75	0.455

Table 5.4 *IGF2* methylation stratified by sex.
Median methylation between males and females was compared by Mann-Whitney U test. *IGF2* methylation is shown separately for each of the three CpG sites and stratified for sex. IQR - Interquartile range.

No significant differences in *IGF2* DNA methylation were found between the sexes when analysed using the Mann-Whitney U test (z=-0.75, p=0.46) (Figure 5.4).

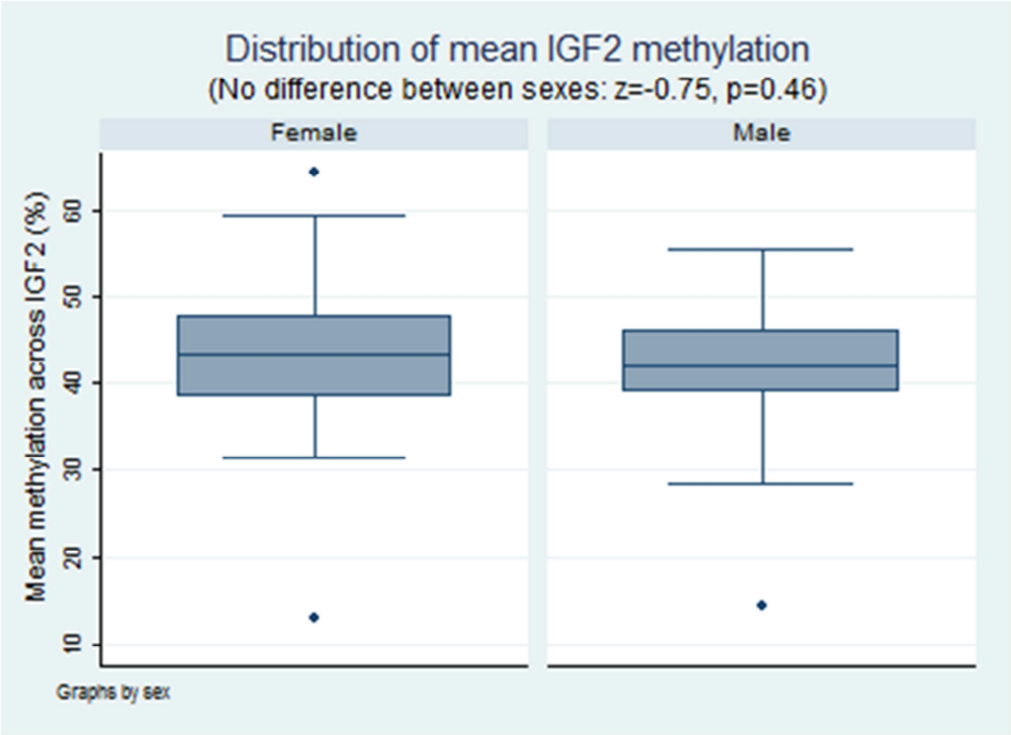


Figure 5.4 Comparison of mean *IGF2* methylation between sexes.
Mean % methylation shown across the *IGF2* gene, box whisker plots shown for each sex. p = 0.46.

The relationship between *IGF2* methylation and age was assessed using Spearman's correlation (Table 5.5). *IGF2* DNA methylation was found to be positively associated with age both at in site one (rho=0.43, p=2.08E-7) and in the mean measure across all three CpG sites (rho=0.37, p=2.00E-5) (Figure 5.5).

		Spearman's Correlation	
***IGF2* CpG Site**	n	rho	p-value
Site 1	132	0.43	2.08E-07
Site 2	132	0.27	0.002
Site 3	132	0.26	0.003
Mean	132	0.37	2.00E-05

Table 5.5. Correlation between *IGF2* methylation and age in children.
Mean age (SD) = 7.35 (0.41) years.

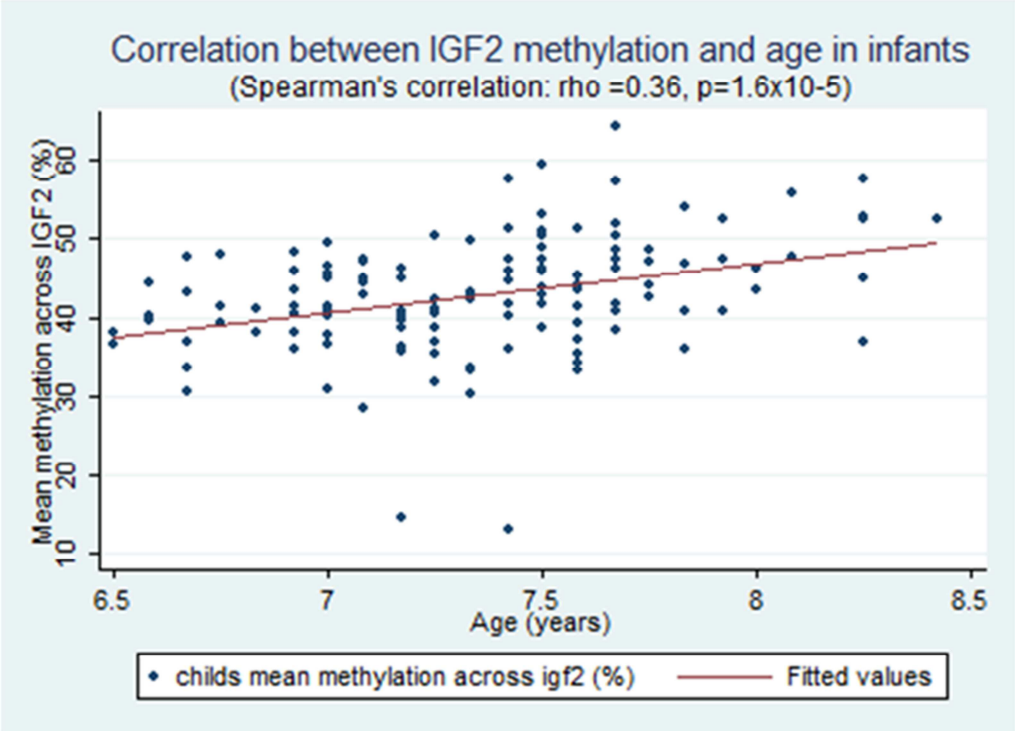


Figure 5.5 Correlation between *IGF2* methylation and age in children.
Mean methylation across the entire *IGF2* gene shown against age of child in years with line of best fit in red.

5.4.3 TACSTD2 methylation

Methylation at seven CpG sites in *TACSTD2* was assessed using Pyrosequencing in children as described above. Correlation between CpG sites in *TACSTD2* was assessed using Spearman's rank correlation. *TACSTD2* showed very strong positive correlation between CpG sites with rho values of greater than 0.90 (Figure 5.6). This again indicated that mean methylation across all three sites could be used in downstream analysis without significant loss of sensitivity as mean methylation across all three sites was a representative measure of the separate methylation values. *TACSTD2* showed a median level of methylation across all three sites of 41.10 (29.48, 49.15) (Table 5.6).

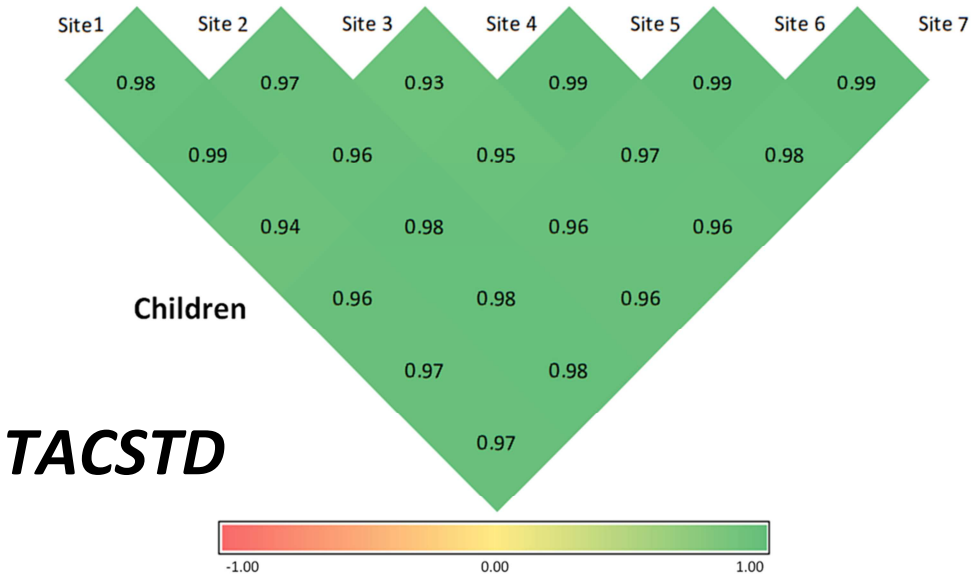


Figure 5.6 Correlation of CpG sites within *TACSTD2* gene.
Degree of correlation assessed by Spearman's rank and shown by colour coding, positive correlation is shown in green.

The relationship between *TACSTD2* DNA methylation and sex was assessed by Mann-Whitney U test (Table 5.6), no significant differences were observed between methylation levels between sexes (Figure 5.7).

	Boys only		Girls only		Normalcy test	Sex comparison	
***TACSTD2* CpG site**	n	Median % Methylation (IQR)	n	Median % Methylation (IQR)	p-value	Stat (z)	p-value
Site 1	45	26.01 (20.25, 34.86)	45	26.90 (19.21, 33.84)	0.010	0.57	0.572
Site 2	45	42.24 (34.02, 51.91)	45	45.28 (30.44, 51.38)	0.002	0.53	0.597
Site 3	45	25.44 (17.17, 32.19)	45	26.75 (16.86, 30.44)	0.002	0.42	0.678
Site 4	45	56.54 (47.01, 66.51)	45	56.61 (41.64, 66.75)	0.001	0.35	0.729
Site 5	45	47.75 (38.70, 58.41)	45	50.00 (35.77, 58.49)	4.00E-04	0.23	0.818
Site 6	45	45.60 (34.16, 52.48)	45	45.69 (30.56, 54.08)	0.002	-0.04	0.968
Site 7	45	34.95 (27.11, 45.18)	45	37.81 (26.91, 44.81)	0.002	-0.07	0.949
Mean	45	39.83 (31.93, 47.38)	45	42.31 (29.43, 49.15)	0.001	0.35	0.726

Table 5.6 *TACSTD2* methylation stratified by sex.
Median percentage methylation was compared between males and females using Mann-Whitney U test. Normalcy was assessed using Shapiro-Wilk normal data test. IQR – Interquartile Range.

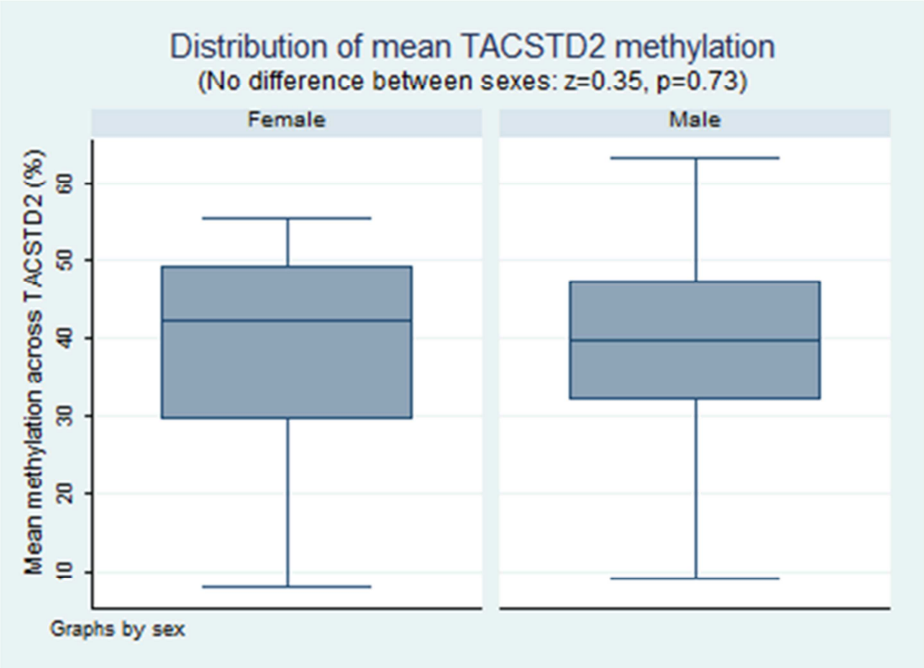


Figure 5.7 Distribution of mean *TACSTD2* methylation stratified by sex.
Mean % methylation shown across the *TACSTD2* gene, box whisker plots shown for each sex. p = 0.73.

The relationship between *TACSTD2* DNA methylation and age was also assessed using Spearman's correlation (Table 5.7). No significant associations were found between *TACSTD2* DNA methylation and age in any CpG site or indeed in the overall mean measure. (Figure 5.8).

		Spearman's Correlation	
***TACSTD2* CpG site**	n	rho	p-value
Site 1	90	-0.07	0.498
Site 2	90	-0.07	0.522
Site 3	90	-0.05	0.611
Site 4	90	-0.05	0.654
Site 5	90	-0.05	0.649
Site 6	90	-0.05	0.605
Site 7	90	-0.03	0.750
Mean	90	-0.05	0.616

Table 5.7 Correlation between *TACSTD2* DNA methylation and age in children.
Mean age (SD) = 7.35 (0.41) years.

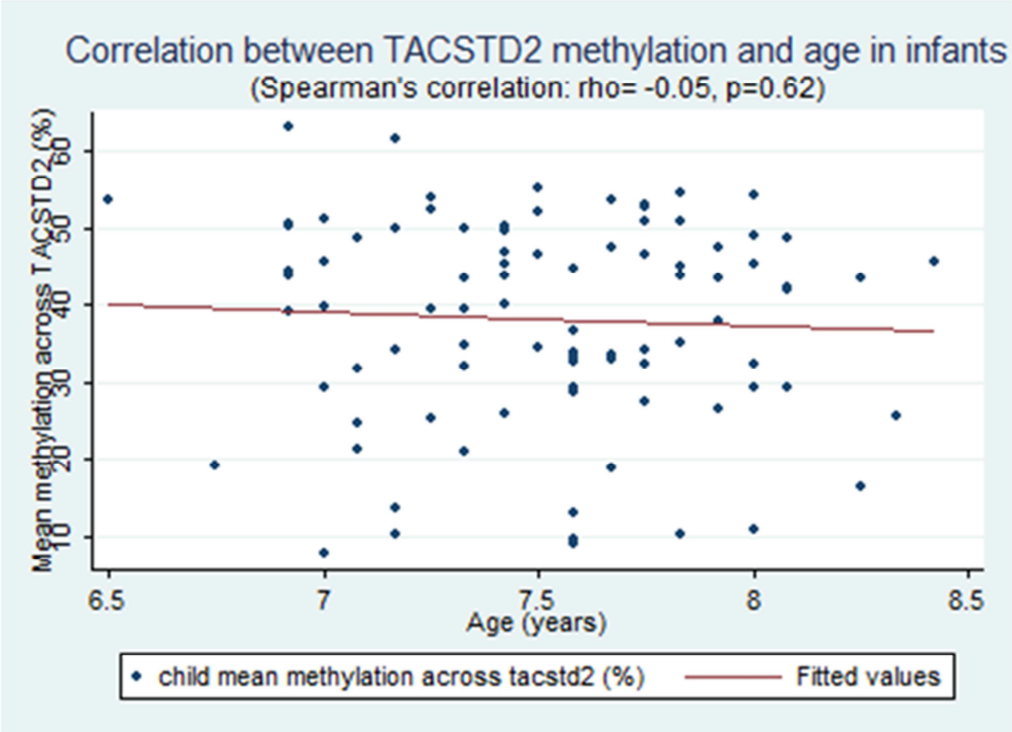


Figure 5.8 Correlation between *TACSTD2* DNA methylation and age in children.
Mean methylation across the entire *TACSTD2* gene shown against age of child in years with line of best fit in red.

5.4.4 Regression analysis

5.4.4.1 *IGF2*

The relationship between *IGF2* DNA methylation and the phenotypic traits bioimpedance and BMI was assessed using a Spearman's correlation stratifying for sex. No significant association between *IGF2* DNA methylation and either of the phenotypic traits was observed, with a p-value of 0.224 in the case of bioimpedance and 0.552 in the case of BMI when the both sexes combined were assessed. (Table 5.8).

Phenotypic measure	Group	n	rho	p-value
Bioimpedance	All	132	-0.11	0.224
	Boys	68	-0.19	0.127
	Girls	64	-0.07	0.608
BMI	All	131	0.05	0.552
	Boys	67	0.04	0.775
	Girls	64	0.07	0.563

Table 5.8 Spearman's correlation of mean *IGF2* methylation against phenotypic traits. (Univariate)

The association between *IGF2* DNA methylation and the phenotypic traits bioimpedance and BMI were also analysed by multiple regression adjusting for age and sex. No significant associations were found with either bioimpedance (p=0.643) or BMI (p=0.456) (Table 5.9).

Phenotypic measure	Coeffecient [95% CI]	p-value
Bioimpedance	-0.39 [-2.06, 1.28]	0.643
BMI	0.02 [-0.03, 0.08]	0.456

Table 5.9 Multiple regression of mean *IGF2* DNA methylation against phenotypic traits. Adjusted for age and sex

5.4.4.2 TACSTD2

The relationship between *TACSTD2* DNA methylation and the phenotypic traits bioimpedance and BMI was assessed using a Spearman's correlation stratifying for sex. No significant association between *TACSTD2* DNA methylation and either of the phenotypic traits was observed, with a p-value of 0.896 in the case of bioimpedance and 0.584 in the case of BMI when the both sexes combined were assessed. (Table 5.10).

Phenotypic measure	Group	n	rho	p
Bioimpedance	All	89	-0.01	0.896
	Boys	45	-0.09	0.555
	Girls	44	0.10	0.525
BMI	All	90	-0.06	0.584
	Boys	45	0.08	0.602
	Girls	45	-0.19	0.220

Table 5.10 Spearman's correlation of mean *TACSTD2* methylation against phenotypic traits. (univariate)

The association between *TACSTD2* DNA methylation and the phenotypic traits bioimpedance and BMI were also analysed by multiple regression adjusting for age and sex. No significant associations were found with either bioimpedance (p=0.834) or BMI (p=0.952) (Table 5.11).

Phenotypic measure	Coefficient [95% CI]	p-value
Bioimpedance	-0.15 [-1.58, 1.28]	0.834
BMI	2.00E-3 [-0.06, 0.07]	0.952

Table 5.11 Multiple regression of mean *IGF2* DNA methylation against phenotypic traits. Adjusted for age and sex

5.5 Discussion

In this chapter gene specific DNA methylation was assessed in the promoter regions of the genes *IGF2* and *TACSTD2*. Mean *IGF2* DNA methylation was positively associated with age when assessed by Spearman's correlation (rho=0.37, p=2.00E-5) (Figure 5.5). The strongest link between differential IGF2 DNA methylation and age has been made in colon cancer where a switch from monoallelic to biallelic imprinting was associated with increasing age (Issa et al., 1996). Previous studies have also reported age-related changes in DNA methylation at the *IGF2* locus, albeit at different CpG sites within the *IGF2* DMR. One such study of individuals aged 58 years (SD=0.35), reported a 3.6% decrease in *IGF2* methylation per decade and a mean *IGF2* methylation of 51% (SD=4.50) (Heijmans et al., 2008). When compared to the *IGF2* methylation level of 42.72% measured in children aged 7.4 years, this suggests that *IGF2* DNA methylation may indeed increase with age. There have been some recent human studies into *IGF2* methylation in which age was used as a covariate in the analysis (Steegers-Theunissen et al., 2009) however this particular study found no association between age and *IGF2* methylation. Another paper assessed IGF2 DMR methylation in human twins and found that the influence of environmental and stochastic factors was the same at age 16.7 years (SD, 2.0) when compared to 44.8 years (SD, 6.8), suggesting a limited role for age-related degeneration of methylation patterns at the *IGF2* locus (Heijmans et al., 2007). The lifecourse trajectory of *IGF2* DNA methylation is difficult to predict because it is an imprinted gene and as such has an extra layer of complication to the maintenance of DNA methylation levels (Cui et al., 2003). However given current theory DNA methylation at this locus could be expected to increase in variability with age as the maintenance of methylation marks by the DNMTs gradually decreases (Liu et al., 2011). *IGF2* promoter DNA methylation may be subject to variability in earlier life however due to environmental and stochastic factors and these could be the source of the association with age in this chapter.

This chapter found no association between mean *TACSTD2* DNA methylation and any of the covariates assessed including age, bioimpedance and BMI. This is interesting given the recent finding in our group that in preterm children increased levels of *TACSTD2* methylation were associated with lower fat mass at age 12 years (Groom et al., 2010). There is very little literature on the relationship between *TACSTD2* DNA methylation and aging, however aberrant *TACSTD2* methylation has been previously implicated in a number of cancers including glioma (Kim et al., 2006) and prostate cancer (Ibragimova et al., 2010). A

recent study also found that *TACSTD2* was one of the top gene hits in the American population with regards to NAFLD development, which is interesting given the well known association between NAFLD and obesity (Angulo, 2007).

There is a broader literature linking age to genome-wide changes in DNA methylation. This includes a study of chronological age from one to 102 years that highlighted a shift towards increased DNA methylation at multiple loci across the genome with increasing chronological age (Hernandez et al., 2011). This study is particularly interesting as it was performed on brain tissue. This is relevant as brain is a post-mitotic tissue and as such changes in DNA methylation might be expected to accumulate over time. This is in contrast to other tissues types from which DNA is regularly extracted such as blood and buccal swabs, where cells are relatively rapidly turned over and as such may not have the same opportunity to accumulate such widespread DNA methylation changes with age (Sugawara et al., 2011). Recent studies have also shown both increased promoter DNA methylation and decreased mRNA expression of the type 2 diabetes modulating gene COX7 in the skeletal muscle of elderly twins (Ronn et al., 2008). Indeed, the role of epigenetic changes in older age is well established (Calvanese et al., 2009).

It is clear both from the results of this chapter and previous literature that DNA methylation is subject to variance with aging in a gene specific manner. The factors which influence these changes during the lifecourse are well characterised and include aberrant maintenance of DNA methylation marks by DNMTs as previously discussed (Richardson, 2003). However, these underlying age-related changes in DNA methylation are less well known. In the case of gene specific promoter methylation changes, it is possible that methylation changes simply track changes in gene expression at individual ages. Alternatively, DNA methylation may be initially independent of gene expression changes and may contribute or be accelerated by gene expression changes, leading to hypomethylation. An alternative hypothesis suggests that methylation is initially attracted to DNA by short stretches of DNA specifically targeted to DNA methylation. These include B1 elements in mouse and Alu elements in human DNA. Once established, DNA methylation could spread in cis with cell replication, resulting in progressive hypermethylation (Fraga and Esteller, 2007).

The findings of this chapter could be build on by replicating this methodology in longitudinal studies where DNA samples could be taken at multiple time points to allow the temporality of any changes in DNA methylation levels to be assessed (Wong et al., 2010). Including a number of different tissue types such as liver, adipose tissue and muscle

(Chen et al., 2010) would further strengthen this analysis especially given the previously mentioned observation that post mitotic tissues such as brain are likely to harbour more persistent changes in DNA methylation (Ma et al., 2011).

The data in this chapter provide no evidence to suggest that differential methylation at the *neither IGF2 nor TACSTD* locus is implicated in the programming of disease risk in children. Although epigenetic plasticity in early life is known to exert an influence on later health (Hochberg et al., 2011) it is likely that epigenetic effects are gene specific. Although we report no association between DNA methylation and health outcomes the assays were locus specific and this does not preclude other loci from being of potential interest in the future. As such further analysis could include a larger number of loci so assess if DNA methylation alters with age in a gene-specific manner (Maegawa et al., 2010).

Chapter 6: The Relationship Between Cardiometabolic Health and DNA Methylation at Age 50

6.1 Background

DNA Methylation is the covalent modification of cytosine residues in the DNA sequence through the addition of a methyl group. This converts cytosine to an alternative nucleotide base, 5-methyl cytosine (5meC) (Rakyan et al., 2001). In the human genome, this covalent modification takes place on cytosine residues that are located 5' adjacent to guanine residues. These sequences of nucleotides are known as CpG sites and they tend to cluster together, forming motifs known as CpG Islands (CGIs). These CGIs make up a maximum of 2% of the genome, and are for the most part un-methylated. Some CGIs however are more highly methylated and these tend to be related to imprinted genes or to the presence of transposons (Waterland and Jirtle, 2003). More recently, another methylation mark on DNA, 5-hydroxymethyl-2'deoxycytidine (5hmC) has been described (Kriaucionis and Heintz, 2009). This is a residue formed when a hydroxy group is added to 5-methylcytosine, catalysed by Methylcytosine dioxygenase (TET1) (Tahiliani et al., 2009). Research into the implications of 5hmC is in its infancy, however it has recently been characterised in mouse, bovine and rabbit zygotes and may play a role in DNA methylation reprogramming (Wossidlo et al., 2011).

CGIs are understood to be important in terms of gene regulation due to their abundance in regions harbouring transcription start sites. Predominantly, a higher level of DNA methylation (that is a greater proportion of cytosine residues that are methylated) in gene regulatory domains such as promoter sequences leads to transcriptional repression. DNA methylation on a genome-wide level can be assessed by utilising repeat interspersed regions such as Alu or LINE-1. Comprising approximately 17% of the human genome, LINE-1, or Long Interspersed Nuclear Element 1 is the most abundant family of non-LTR retrotransposons found in the genome (Belancio et al., 2010). LINE-1 is considered a potential mutagen as its transposition can induce DNA strand breaks and genomic instability (Belgnaoui et al., 2006). Whilst repetitive elements such as LINE-1 are heavily methylated in normal tissues, they are hypomethylated in cancers (Ogino et al., 2008). A recent study has reported that LINE-1 methylation levels, when considered together with exposure to dietary folate and alcohol consumption, predict colon cancer risk (Schernhammer et al., 2010). Although most closely related to cancer outcomes LINE-1

methylation has more recently been linked with a diverse range of disease outcomes. In the Boston-based Normative Aging Study for example, individuals suffering from prevalent ischemic heart disease (IHD) and stroke exhibited decreased levels of LINE-1 methylation. Individuals in this cohort with decreased levels of LINE-1 methylation were also shown to be at higher risk for incident IHD, stroke and total mortality (Baccarelli et al., 2010b). Global DNA methylation has also been linked to Alzheimer's Disease (AD) in a recent study which showed individuals exhibiting decreased performance in mental tasks had lower LINE-1 DNA methylation (Bollati et al., 2011). Other pathologies recently associated with LINE-1 DNA methylation include Systemic lupus erythematosus (Lupus), (Nakkuntod et al., 2011), Pre-eclampsia (Gao et al., 2011a) and coronary heart disease (Chowdhury et al., 2011).

Studies both in experimental animals and in human cohorts have shown that environmental insults can influence DNA methylation (Mathers et al., 2010). These include exposure to perfluorooctane sulfonate (PFOS), an environmental pollutant (Wan et al., 2010), prenatal tobacco smoke exposure, polycyclic aromatic hydrocarbons (PAHs) found in cigarette smoke condensate (Liu et al., 2010a), biomarkers of lead levels in both adults (Wright et al., 2010) and cord blood (Pilsner et al., 2009b), traffic particulates (Baccarelli et al., 2009) and plasma homocysteine (Fryer et al., 2011).

Aside from more extreme environmental insults a number of recent studies in the North Texas Healthy Heart Study have shown that milder lifecourse exposures may influence levels of LINE-1 global DNA methylation. A 'healthy' dietary pattern showed a dose dependent association with DNA hypomethylation (Zhang et al., 2011b) whereas individuals with higher levels of physical activity exhibited increased levels of LINE-1 methylation (Zhang et al., 2011a).

Given the increasing evidence in current literature linking global DNA methylation with disease phenotypes or phenotypic traits that could be indicative of early stage disease other than cancer, the present study addresses the hypothesis that genome-wide DNA methylation, measured at age 50 when disease related phenotypic traits are discernable, is associated with traits indicative of early stage metabolic disease.

6.2 Aims and objectives

The aim of this section of study was to assess if LINE-1 global DNA methylation was associated with markers of metabolic health in adults aged 50 years. To this end single base resolution sequencing of LINE-1 repeat elements was undertaken as these elements act as a surrogate for global DNA methylation levels. These global methylation measurements were then regressed against cardiometabolic traits in the individuals to assess any possible associations.

6.3 Methods

6.3.1 Study participants

The Newcastle Thousand Families birth cohort consists of 1142 individuals born in May and June 1947 to mothers resident within the city of Newcastle upon Tyne in northern England (Pearce et al., 2009). Two thirds of these children were followed up until the age of 15 years, with detailed information collected prospectively on their health, growth and socio-economic circumstances. Participants in this investigation were cohort members who either contacted the study team in response to media publicity or were traced through the National Health Service Central Register during the 1990's. Between October 1996 and December 1998 self-completion questionnaires on health and lifestyle were sent out and members invited to attend for clinical examination. Of the surviving 89% traced, 574 returned lengthy questionnaires detailing their family history and lifestyle and 412 attended clinical examinations which involved giving blood to be used in DNA analysis (Pearce et al., 2009).

6.3.2 Clinical assessments of outcomes and adult height and weight at age 49-51 years

Assessments were performed in the morning following an overnight fast. Height and weight were measured and body mass index (BMI) was calculated. Waist and hip circumferences were measured according to the protocol of the World Health Organisation Monitoring Trends and Determinants in Cardiovascular Disease (MONICA) project (WHO, 1990). Percent body fat was estimated from impedance measured using a Holtain body composition analyser (Holtain Ltd, Crymych, Wales, UK). All lipid analyses were performed on a DAX analyser (Bayer, Basingstoke). Total cholesterol was measured using

a cholesterol oxidase/peroxidise method with calibrants traceable to the Centres for Disease Control definitive method. Serum HDL cholesterol was measured using a cholesterol oxidase method after precipitation of apolipoprotein B containing lipoproteins with phosphotungstic acid and magnesium chloride (inter-assay coefficient of variation 2.2%). LDL cholesterol levels were derived by the Friedewald method (Friedewald et al., 1972) and the HDL:LDL ratio was calculated. Triglyceride concentrations, excluding glycerol, were estimated by a lipase-glycerol kinase method. Serum insulin levels at 0, 30 and 120 minutes were determined by ELISA (Dako Ltd, Ely, UK) (interassay coefficients of variation 3.1% and 3.3% respectively) (Pearce et al., 2006). Plasma glucose concentrations at 0, 30 and 120 minutes (after a 75g oral glucose load) were measured on a Yellow Springs Analyser (YSI Stat Plus 2300; Yellow Springs Instruments, Farnborough, UK).

6.3.3 Measurement of global DNA methylation

DNA was extracted from peripheral blood samples using a Nucleon BACC2 kit (Tepnel Life Sciences, UK). 1μg of DNA sample was bisulphite modified using the Zymo EZ DNA Methylation Gold kit (Cambridge Bioscience, Cambridge) using the manufacturers' standard protocol. 1μg of bisulphite modified DNA was PCR amplified using 2 x HotstarTAQ Mastermix (Qiagen), 2mM MgCl2 (Qiagen) and 0.2μM of each primer (LINE-1 Forward Primer - 5' – TTT TGA GTT AGG TGT GGG ATA TA – 3' and LINE-1 Reverse Primer – BIO-5' – AAA ATC AAA AAA TTC CCT TTC – 3'). PCR conditions were as follows: 95˚C for 15 minutes, 50 cycles of 95˚C for 15 seconds, 60˚C for 30 seconds and 72˚C for 15 seconds and finally 72˚C for five minutes. 5μL of amplicons were utilised for downstream single strand preparation and hybridisation of 0.5μM sequencing primer (5'-GGG TGG GAG TGA T-3'), using a vacuum prep tool and workstation according to manufacturer's instructions (Qiagen). Samples were then run on a PyroMark MD Pyrosequencer (Qiagen, UK) using an assay designed and optimised for LINE-1. The dispensation sequence for LINE-1 was: GTCGATTAGTAGTCTGTCGCTC.

6.3.4 Statistical analysis

LINE-1 Pyrosequencing generated a percentage measure of methylation at each of three CpG sites. Samples were run in duplicate (post-bisulphite modification) and a mean percentage methylation value calculated for duplicates for each of the three CpG sites. Correlation between all three CpG sites was high ($p<0.001$); therefore a composite mean of all three sites was calculated and tested for association with outcome variables. DNA methylation showed a skewed distribution (tested using a Shapiro Wilk test) so values were log transformed before further analysis. Linear regression was used to describe the relationship between anthropometric measures and blood biomarkers (the dependent variables) and the independent variable, log-transformed global DNA methylation at age 50 years. Regression coefficients (b) and corresponding 95% confidence intervals are reported showing the unit change in log methylation per unit increase in each dependent variable after adjustment for sex. Overall R^2 values for the models including both log-transformed methylation are given as percentages. Direct R^2 values for methylation were estimated by subtracting the R^2 value for the models only including sex from the corresponding models including both log-transformed methylation and sex. This gives an estimate of the direct association between outcome and methylation after adjusting for any potential mediation through sex.

6.4 Results

6.4.1 Cohort details and sample selection

A total of 228 individuals (85 males and 143 females) from the Newcastle Thousand Families Study (NTFS) were analysed with respect to LINE-1 methylation. The individuals were selected from the full NTFS cohort of 1029 individuals initially recruited to the study on two criteria. Firstly those individuals who attended clinical follow-up at age 49-51 years and secondly, on the availability of DNA of sufficient quantity and quality to undertake DNA methylation analysis. Descriptive data for all variables used in the current analysis are given in Table 6.1. Significant differences in all but total cholesterol levels were seen between males and females.

	Males		**Females**		**Normalcy test**	**Sex comparison**	
Phenotypic variable	n	Median (IQR)	n	Median (IQR)	p-value	z	p-value
BMI (kg/m^2)	85	26.85 (24.38, 29.44)	143	25.00 (22.46, 27.91)	1.71E-8	2.64	0.008
Waist/hip ratio	85	0.96 (0.92, 0.99)	143	0.79 (0.75, 0.83)	1.25E-6	11.8	2.48E-32
Body fat (%)	85	38.30 (31.80, 42.00)	141	42.10 (35.20, 47.20)	0.047	-3.88	1.06E-04
Fasting glucose (mmol/l)	85	5.43 (5.09, 5.74)	143	5.03 (4.78, 5.33)	3.67E-25	5.77	8.15E-09
Total cholesterol (mmol/l)	85	5.33 (4.49, 6.08)	143	4.90 (4.28, 5.76)	7.10E-4	1.6	0.109
Total triglycerides (mmol/l)	85	1.16 (0.83, 1.79)	143	0.87 (0.63, 1.31)	1.84E-20	3.62	3.00E-04
HDL cholesterol (mmol/l)	85	1.00 (0.83, 1.19)	143	1.13 (0.85, 1.39)	9.63E-10	2.97	0.003
LDL cholesterol (mmol/l)	85	4.00 (3.2, 4.71)	143	3.65 (2.90, 4.50)	0.079	2.09	0.037
HDL:LDL ratio	85	0.24 (0.19, 0.32)	143	0.29 (0.22, 0.45)	4.37E-21	-2.96	0.003

Table 6.1 Phenotypic variables measured at age 50 stratified for sex.
IQR – Interquartile range; sex comparison by Mann-Whitney U test.

6.4.2 Quantification of LINE-1 DNA methylation

LINE-1 DNA methylation was successfully quantified in all 229 samples. The assay provides quantitative assessment of three CpG sites that result from arbitrary priming of LINE-1 repetitive elements across the genome. Upon visual and formal statistical inspection, the distribution of methylation measures was deviated from a normal distribution (Shapiro Wilk p-value of 4.29E-10) (Figure 6.1). The median level of methylation across all individuals was 52.74 (IQR: 51.51, 54.89) (Table 6.2). There was no significant difference in levels of LINE-1 methylation between males and females (Figure 6.2).

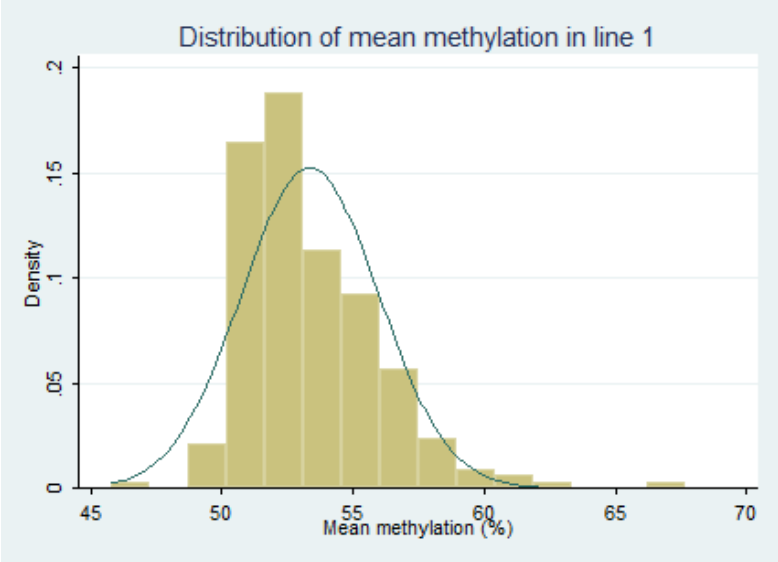


Figure 6.1 Histogram of the distribution of mean LINE-1 DNA methylation throughout the samples.

	Males		Females		Normalcy test		Sex comparison	
LINE-1 CpG site	n	Median (IQR)	n	Median (IQR)	z	p-value	z	p-value
Site 1	85	72.19 (70.42, 74.28)	144	72.08 (70.64, 74.47)	4.06	2.45E-05	-0.36	0.721
Site 2	85	36.80 (34.79, 38.41)	144	36.26 (35.03, 37.99)	6.16	3.57E-10	1.09	0.277
Site 3	85	50.61 (48.89, 52.91)	144	50.08 (48.56, 52.14)	5.87	2.24E-09	1.17	0.241
Mean	85	52.92 (51.55, 55.18)	144	52.53 (51.42, 54.80)	6.13	4.29E-10	0.65	0.515

Table 6.2 LINE-1 DNA methylation across all three CpG sites stratified by sex.
IQR – Interquartile range; sex comparison by Mann-Whitney U test. Percentage methylation is shown for all three CpG sites within the LINE-1 amplicon on the mean calculated over all three sites.

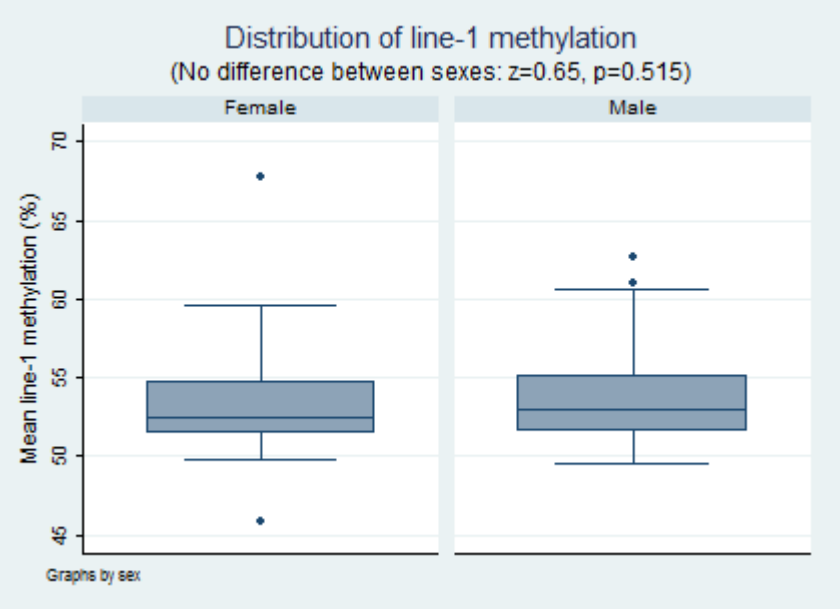


Figure 6.2 Relationship between LINE-1 DNA methylation and sex.
Distribution of LINE-1 methylation between sexes shown by box whisker plot. Sex comparison by Mann-Whitey U test.

DNA methylation between the three CpG sites in the LINE-1 locus showed a variance of between 36.26% and 72.19%. All three CpG sites were highly positively correlated, with a Spearman's rho value of greater than 0.60 for all sites (Figure 6.3). This indicated that Mean methylation calculated across all three CpG sites could be used for the association analysis as a surrogate for all three CpG sites.

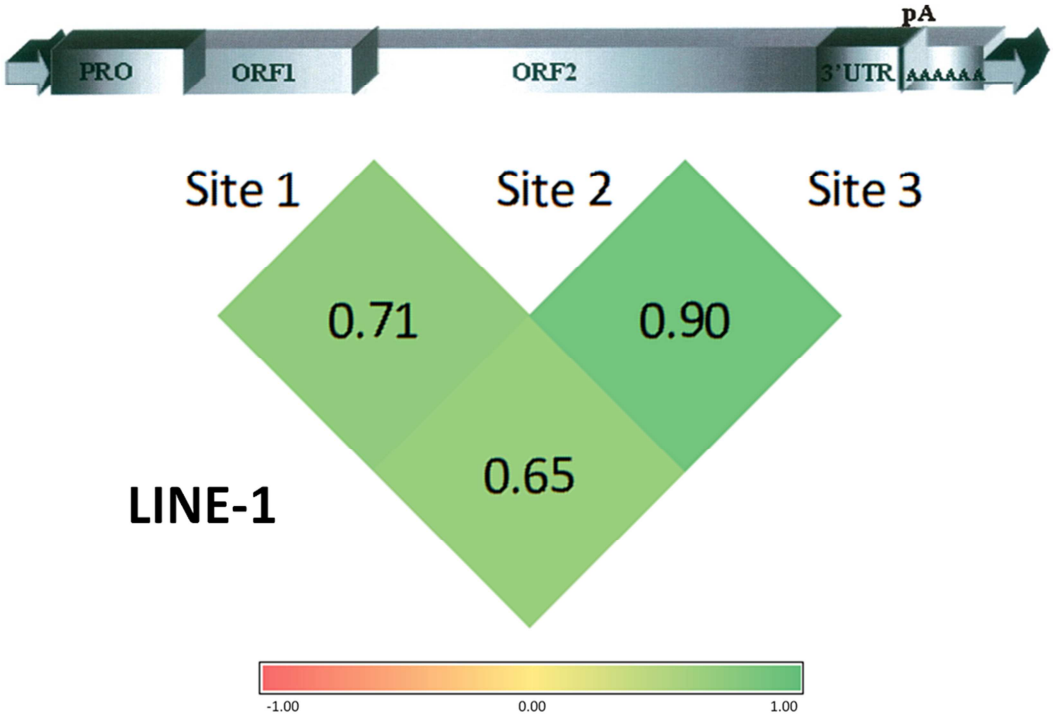


Figure 6.3 Correlation of DNA methylation between CpG sites in LINE-1.
Correlation between all three CpG sites in LINE-1 shown with colour coding. The higher the correlation, the closer to 1.0.

6.4.3 Association analyses between LINE-1 methylation and metabolic health phenotypes

Multiple regression analyses, with adjustment for sex, were performed between LINE-1 methylation and metabolic health phenotypes. Results are given in Table 6.3. Increased LINE-1 methylation was associated with increased fasting glucose (p=0.02), total cholesterol (p=0.005), total triglycerides (p=0.003) and LDL cholesterol (p=0.001), and with decreasing HDL cholesterol (p=0.003) and the HDL:LDL ratio (p=0.003). As can be seen from the R^2 values in Table 6.3, when sex was adjusted for LINE-1 methylation accounted for 5-10% of the variability in outcome measures.

Variable	Coefficient [95% CI]	p-value	R^2 (%)
BMI (kg/m^2)	0.04 [-0.18, 0.26]	0.721	0.94
Waist/hip ratio	2.00E-3 [2.00E-4, 5.00E-3]	0.072	65.42
Body fat (%)	0.07 [-0.37, 0.51]	0.756	5.61
Fasting glucose (mmol/l)	0.05 [0.007, 0.09]	0.024	9.97
Total cholesterol (mmol/l)	0.09 [0.03, 0.15]	0.005	5.05
Total triglycerides (mmol/l)	0.07 [0.02, 0.11]	0.003	9.68
HDL cholesterol (mmol/l)	-0.03 [-0.04, 8.00E-3]	0.004	8.91
LDL cholesterol (mmol/l)	0.10 [0.04, 0.16]	0.001	7.07
HDL:LDL ratio	-0.02 [-0.03, 6.00E-3]	0.003	6.96

Table 6.3 Regression analysis of phenotypic variables against LINE-1 DNA methylation. Adjusted for sex.

Similar patterns of association were also observed for all but fasting glucose following non-parametric univariate correlation analyses within the complete cohort (Table 6.4). In addition, correlations between LINE-1 methylation and serum measures were also apparent upon sex stratification although not all observations reached statistical significance across these smaller subgroups (Table 6.4).

	All			Male			Female		
Variable	n	rho	p-value	n	rho	p-value	n	rho	p-value
BMI (kg/m^2)	228	0.01	0.768	85	0.08	0.476	143	-0.04	0.642
Waist/hip ratio	228	0.11	0.101	85	0.07	0.541	143	0.16	0.051
Body fat (%)	226	0.04	0.534	85	0.02	0.863	141	0.08	0.374
Fasting glucose (mmol/l)	228	-0.004	0.95	85	0.09	0.411	143	-0.06	0.466
Total cholesterol (mmol/l)	228	0.13	0.058	85	0.19	0.089	143	0.08	0.343
Total triglycerides (mmol/l)	228	0.14	0.032	85	0.17	0.118	143	0.13	0.132
HDL cholesterol (mmol/l)	228	-0.18	0.006	85	-0.18	0.100	143	-0.19	0.026
LDL cholesterol (mmol/l)	228	0.16	0.015	85	0.18	0.091	143	0.14	0.091
HDL:LDL ratio	228	-0.21	0.002	85	-0.24	0.025	143	-0.19	0.023

Table 6.4 Non- parametric correlation of phenotypic variables against LINE-1 DNA methylation.
Stratified for sex.

6.5 Discussion

The data presented in this chapter show that increased LINE-1 DNA methylation is associated with a number of blood based biomarkers of metabolic health and provides some of the first evidence of an association between LINE-1 methylation and phenotypic traits other than cancer. There is widely documented evidence that LINE-1 methylation is modulated by environmental exposures TVR (Wan et al., 2010, Breton et al., 2009, Liu et al., 2010a, Wright et al., 2010, Pilsner et al., 2009b, Baccarelli et al., 2009, Fryer et al., 2011) and the current study suggests that DNA methylation may provide a mechanistic link between these environmental exposures and the development of disease related traits. There is considerable interest in the role of epigenetic mechanisms in common complex disease (Feinberg, 2008), particularly those with a prominent environmental component. It is likely that epigenetic factors contribute to the inter-individual differences in response to environmental exposures (Mathers, 2008) and to the pathogenesis of such diseases (Turan et al., 2010).

The results show a number of associations between LINE-1 methylation and blood-based biomarkers, all of which are risk factors for cardiovascular disease and/or type 2 diabetes. High levels of fasting glucose are associated with the development of both cardiovascular disease and diabetes (Turan et al., 2010). A positive association was observed between LINE-1 methylation and total cholesterol and triglycerides. Elevated concentrations of these blood lipid markers are strongly associated with increased risk of cardiovascular disease (Sarwar et al., 2010). Further, the present results show that LINE-1 DNA methylation was associated with both increased LDL-cholesterol and decreased HDL-cholesterol concentrations. These changes in opposing directions are those expected in individuals at increased cardiovascular disease risk (Barter et al., 2007, Mertens and Holvoet, 2001). Taken together, these adverse changes in both blood glucose and blood lipid status may be indicative of early stage cardiovascular disease. As these data were collected in individuals at age 50 with no evidence of overt cardiovascular disease it was not possible to explore a potential association between LINE-1 methylation and subsequent disease phenotype. Indeed, this could be perceived as a strength of this study as it removes the potential confounding effect of disease status on DNA methylation patterns. The cohort is being followed up longitudinally which will provide the opportunity to ascertain the predictive utility of LINE-1 methylation at age 50 for later disease risk.

These correlative findings are some of the first observations to link DNA methylation levels with disease-related traits. They suggest that DNA methylation – and epigenetic mechanisms more widely - might be important in determining risk of common complex diseases such as type 2 diabetes and cardiovascular disease. However, it will be crucial to understand whether DNA methylation is causal in altering blood-based biomarkers such as fasting glucose and lipid levels or whether the reverse is the case. There is limited evidence to suggest that altering glucose concentrations changes DNA methylation patterns (Sharma et al., 2008), but to our knowledge no direct evidence to link lipid concentrations to perturbed DNA methylation, or vice versa. A recent study of genome-wide methylation in cord blood DNA highlighted numerous methylation-variable loci whose biological roles were related to lipid metabolism, suggesting a causal influence of altered methylation on lipid levels (Fryer et al., 2011). A previous study of patients with coronary artery disease and controls showed that global DNA methylation was associated with coronary artery disease risk, and that this association was accentuated by increased plasma homocysteine concentration. (Li et al., 2010c). One method that could provide further insight into the direction of causality is the adoption of a Mendelian randomization approach, as proposed recently by Relton and Davey Smith (Relton and Smith, 2010). This approach, termed 'genetical epigenomics', involves the use of genetic variants as proxies for specific exposures, such that an association between genotype and DNA methylation would be indicative of a causal relationship (as lipid levels could not plausibly influence genotype and thus the possibility of reverse causation is removed). Numerous genetic variants have recently been reported to influence blood lipid profiles (Inouye et al., 2010) and these could collectively be used as a proxy for lipid levels, to test the association between lipid levels and DNA methylation. Additional approaches to explore the causal relationship could include in vitro studies where glucose and lipid concentrations are manipulated and DNA methylation levels measured or the analysis of serial samples collected longitudinally from the same individuals where temporal changes in both methylation patterns and metabolic biomarkers can be delineated.

The observations reported relate to global DNA methylation measured using the LINE-1 assay, which measures cytosine methylation in non-coding sequence randomly primed across the genome. The functional consequences of DNA methylation at these sites with regard to those specific disease traits considered remains unknown and thus LINE-1 can merely act as a representative biomarker of overall methylation status (global DNA methylation). Interrogation of gene-specific methylation in genes and pathways directly implicated in lipid metabolism and glucose homeostasis may provide greater insight. Of

relevance is a recent study reporting correlation between LINE-1 methylation and methylation of gene-specific CpG sites as measured using the Illumina 27K human methylation array, indicating that LINE-1 may also be representative of genome-wide methylation of gene regulatory regions (Fryer et al., 2011).

In summary, the evidence presented supports the hypothesis that global DNA methylation at age 50 years is associated with biomarkers of metabolic health. These cross-sectional associations do not allow conclusions to be drawn with respect to the direction of causation. However the link between methylation and biomarkers of metabolic health remains highly plausible and may have important implications for prediction, early diagnosis and prevention of common complex diseases such as cardiovascular disease and type 2 diabetes (Relton and Smith, 2010).

A manuscript based on the contents of this chapter has been published in the International Journal of Epidemiology (Global LINE-1 DNA methylation is associated with blood glycaemic and lipid profiles - Mark S. Pearce, James C. McConnell, Laura M. Barrett, Louise Parker, John C Mathers, Caroline L. Relton) and is included in Appendix 1. The manuscript was drafted, reviewed and revised with contributions from the above authors.

Chapter 7: Early Life Influences and their Effect on DNA Methylation Patterns in Later Life

7.1 Background

In addition to the large body of evidence linking low birth weight to subsequent adverse health outcomes, there is increasing data to support a link between premature birth and intrauterine growth restriction to the development of a metabolic disease-like phenotype in later life (Stettler and Iotova, 2010, Catalano et al., 2009, Kaijser et al., 2009). One epidemiological study has indicated that individuals born prematurely had a greater risk of developing type 2 diabetes and the associated seqeulae of obesity (Hofman et al., 2004). In animal models it has also been shown that intrauterine growth restriction and associated catch-up growth led to an increased risk of obesity in later life (Shahkhalili et al., 2010). Another recent study in humans demonstrated that low birth weight is associated with obesity risk at age six years (Taveras et al., 2009). Adults with low birth weight have also been shown to perform less exercise than their normally weighted peers, suggesting a role for early life programming of energy expenditure leading to a greater susceptibility to obesity (Kajantie et al., 2010). These and other similar observations have led to the investigation of potential mechanisms that may link early life exposures with later phenotypic traits including obesity. As alluded to throughout this thesis, one such candidate mechanism is via epigenetic alterations which might be induced in early life and persist across the lifecourse.

There is an increasing body of literature to support the hypothesis that early life exposures, including nutritional factors, influence DNA methylation patterns in later life (Murgatroyd et al., 2009, Gluckman et al., 2008). This evidence does however often rely on small sample sizes and the analysis of only limited DNA methylation measures. A definitive study has yet to be completed nonetheless there is considerable enthusiasm in this area of medical research and a strong motivation to contribute to current understanding about the potential role of epigenetic variation in mediating the effects of early life exposure on later disease risk.

Examples exist that have linked *in utero* and early life exposure to altered methylation patterns either at birth, during childhood or in some instances later in adulthood. Early life exposure to tobacco smoke has been demonstrated to induce differential methylation in both a gene specific and global manner. Prenatal lead exposure has also been shown to

have an inverse relationship with cord blood global DNA methylation (Pilsner et al., 2009b). Maternal exposure to cocaine in rats has been shown to alter global DNA methylation in neonatal offspring (Novikova et al., 2008). Similarly, in humans, exposure to famine during the prenatal period has been shown to exert a gene specific effect on DNA methylation levels (Heijmans et al., 2008), causing hypomethylation at some loci and hypermethylation at others (Tobi et al., 2009). This work is of particular interest due to what would appear to be the persistent nature of epigenetic changes across the lifecourse in the face of an *in utero* exposure. If *in utero* exposure to famine can induce such persistent effects, is it possible that other early life experiences might do the same?

The assessment of the impact of early life influences on DNA methylation in humans raises the question of exposure measurement. In a retrospective cohort study, or even in a prospectively collected longitudinal cohort study we may not have access to the optimal, detailed exposure measures in study participants in early life. Lifecourse studies are often reliant upon historical data that sometimes lacks the precision with which measures might be taken in current research settings. However there are a number of commonly measured variables as well as derived variables that can be utilised in an investigation of this type to act as surrogate measures of early life exposures.

One such measurement is birth weight, which has been used extensively as a surrogate measure of *in utero* growth and as a predictor of health outcomes in adulthood. Indeed low birth weight has been associated with a large number of adult morbidities not confined to obesity (Yu et al., 2011) including type 2 diabetes (Whincup et al., 2008) and cardiovascular disease (Conen et al., 2010). Birth weight although not an early life 'exposure' *per se* acts as surrogate for other early life influences such as maternal nutrition and *in utero* growth (Yajnik and Deshmukh, 2008) or placental insufficiency (Henriksen and Clausen, 2002) and so represents a useful measure of early life influences.

Babies born early usually have a low birth weight. As such birth weight is commonly adjusted for gestational age. Evidence suggests that babies born small for gestational age are predisposed to type 2 diabetes, increased adiposity and other metabolic syndrome type phenotypes in later life. (Ibanez et al., 2006).

TI, a measure of conditional weight gain, was derived by work conducted by Wright et al (Wright et al., 1998, Wright et al., 1994). TI uses the weight of the infant during the early weeks of its life as a baseline with which to compare subsequent weights that have been adjusted for regression to the mean. This requires the transformation of weight measures

into standard deviation scores using a computer algorithm and allows the weight gain of infants who are not following usual growth trajectories to be more effectively monitored. It has been shown to be a useful measure of nutritional growth delay and so is implicated in the development of a metabolic syndrome like phenotype in later life (Gardner et al., 2009).

Breastfeeding is also a useful measure of early life nutritional exposure, and is one of particular interest in the context of the developmental programming of obesity given the postulated protective effect of breastfeeding (Mayer-Davis et al., 2006). There is however little evidence in current literature (largely because of lack of attention not due to null findings) linking duration of breastfeeding with DNA methylation in later life.

Gene specific DNA methylation changes at key loci involved in the insulin signalling pathway may play a role in the programming of birth weight, indeed recent studies have shown differential methylation at the *IGF2* locus in children born small for gestational age (Tobi et al., 2011). Another recent study has shown that new born babies with low or high birth weight had significantly lower LINE-1 global methylation levels compared to normal weight infants after adjusting for gestational age, sex, maternal age at delivery, and maternal smoking (Michels et al., 2011). As such assessing whether these gene specific and global methylation changes persist in later life will help to evaluate if early life influences exert an effect on epigenetic programming.

7.2 Aim and objectives

The aim of this chapter was to test the hypothesis that early life influences affect DNA methylation levels and that these epigenetic changes persist in later life. Measures of early growth and nutrition including gestational age, birth weight and length of breastfeeding were assessed alongside measures of global (LINE-1*)* and gene specific (*IGF2, TACSTD2*) DNA methylation in adults aged 49-51 years.

7.3 Methods

7.3.1 Study participants

Two previously described cohorts were utilised in this portion of the study, the Gateshead Millennium Study (GMS) cohort which is more fully described in Chapter 5 and the Newcastle Thousand Families Study (NTFS) which is described in Chapter 6. A total of 132 individuals from the GMS cohort were assessed for *IGF2* DNA methylation and 90

individuals were assessed for *TACSTD2* DNA methylation. DNA from a total of 229 individuals from the NTFS cohort was analysed for LINE-1 methylation.

7.3.2 Assessment of early life influences and exposures

Early life influences that might plausibly influence DNA methylation patterns in later life were included in the current study. In the GMS, gestational age (days) and birth weight (g) were recorded from delivery records. Birth weight z-score was derived by subtracting mean birth weight for the cohort from measured birth weight for the individual and dividing this result by the standard deviation of the cohort (Parkinson et al., 2011). Thrive index (TI) as a measure of conditional weight gain was derived from algorithms derived by earlier work on this study cohort conducted by Wright (Wright et al., 1998) Information on breastfeeding duration (greater than four months, greater than six weeks, less than six weeks or never) was collected by questionnaire at age eight years. In the NTFS, gestational age (days), birth weight (kg) and duration breastfed (weeks) were abstracted from early life medical records.

7.3.3 Measurement of DNA methylation by pyrosequencing

In the case of the GMS DNA was extracted from saliva using the Oragene Kit (DNA Genotek, UK) whereas in the NTFS DNA was extracted from whole blood using the Qiagen Tissue MiniPrep kit (Qiagen, Crawley, UK). DNA yield ranged from 5-50μg. DNA quality was assessed by NanoDrop spectrophotometer with a 260/230 ratio being indicative of purity. All samples showed a ratio of between 1.6 and 2.0, with these values falling between the minimum values indicating sufficient DNA purity for downstream processing. In both sample series 1μg of genomic DNA was bisulphite modified using the EZ Methylation Gold™ Kit (Zymo, Cambridge, UK) according to the manufacturers' protocol. 1μl of bisulphite modified DNA was then amplified in a PCR reaction containing 12.5μl Hotstar TAQ Mastermix (Qiagen, UK), 2.5μl of 25mM $MgCl_2$ (Qiagen, UK), 3.0μl dH_2O, 0.5μl of 100pmol/μl forward primer and 0.5μl of 100pmol/μl reverse primer (one of which is biotin labelled). The Mastermix excluding $MgCl_2$ was made as follows: 12.5μl Hotstar TAQ Mastermix (Qiagen, UK), 5.0μl dH_2O, 0.5μl forward primer and 0.5μl reverse primer (one of which is biotin labelled). In each Mastermix, 2.5ng DNA was added to each well. Primers are detailed in Table 7.1.

Assay	Forward Primer	Reverse Primer	Sequencing Primer
LINE-1	TTTTGAGTTAGGT GTGGGATA TA	AAAATCAAAAAAT TCCCTTTC	GGGTGGGAGTGAT
IGF2	AGTAAGAAATTGG ATAGG	AAACCCCAACAAA AACCACT	TTTTTTAGGAAGTA TAGTTA
TACSTD2	CTAGGTACTGTAC TGTCA	ACTCACTAGTACG ACAATA	TCTAACCAGGTAA TTGTCCAC

Table 7.1 Forward, reverse and sequencing primers for both assays.
Forward and reverse primers were diluted to a standard concentration of 100pmol/µl.

The PCR reactions were then run on a LabCycler under the following reaction conditions:

95°C for 15 minutes

50 cycles: 95°C for 15 seconds

60°C for 30 seconds

72°C for 15 seconds

72°C for five minutes

4°C ∞

The pyrosequencing PCR samples were then cleaned up to single stranded DNA using the Vacuum Prep Workstation (VPW) (Biotage, UK) as previously described. The pyrosequencing plate was then incubated at 80°C for two minutes and the samples run in duplicate on a Pyromark MD Pyrosequencer (Biotage, UK).

7.3.4 Statistical analysis

Early life influence measurements were dichotomised by sex and the sexes compared by T test, Mann Whitney U test or Chi squared test as appropriate. The measures were assessed for normality using the Shapiro-Wilk test. In the GMS cohort the samples selected were assessed to ascertain if they were representative of the cohort as a whole. Samples not included were compared to samples included by T test or Mann Whitney U test as appropriate. Further, samples that were selected for either or both of the gene specific assays (*IGF2* and *TACSTD2*) were compared to the overall cohort profile to ensure that their phenotypic traits were representative of the whole using T test or Mann Whitney U test as appropriate.

Pyrosequencing generated a percentage measure of methylation at each of the CpG sites within the two gene loci investigated. Similarly a percentage measure of global methylation is generated from the LINE-1 assay that arbitrarily primes repeat regions throughout the genome. All samples were run in duplicate (post-bisulphite modification) and a mean percentage methylation value calculated for duplicates for each of the CpG sites within amplicon. Normality of the DNA methylation data in each of the amplicons was assessed using the Shapiro Wilk test, and sex-dichotomised values were compared using T test or Mann Whitney U test as appropriate.

Linear regression was used to describe the relationship between DNA methylation (the dependent variables) and the independent variable, the early life exposure measurements. Regression coefficients (b) and corresponding 95% confidence intervals are reported showing the percentage change in DNA methylation per unit change in each early life exposure measurement after adjustment for age and sex. Overall R^2 values for the models are given as percentages.

Given the non-normal distributions of methylation and early life measurements, further analysis was performed by Spearman's rank correlation (a non-parametric test) stratifying for sex with breast-feeding analysed by Kruskal–Wallis one-way analysis of variance.

7.4 Results

7.4.1 Representative phenotypic data

Lifecourse exposure variables for individuals utilized were compared to those of individuals from the full GMS cohort to determine whether this subgroup was representative of the whole (Table 7.2). No significant differences were found.

The lifecourse exposure variables in the two gene specific subsets selected from the GMS cohort were also assessed to discern if there were any significant differences between them (Table 7.3). After comparison using T test or Mann-Whitney U test no significant differences between the two subsets were detected.

	Samples not included		**Samples included**		**Normalcy test‡**	**Comparison**	
Variable	n	Median (IQR)	n	Median (IQR)	p-value	stat	p-value
IGF2							
Gestational age (days)	477	40.00 (38.00, 41.00)	132	40.00 (39.00, 41.00)	1.03E-6	-1.61*	0.108*
Birth weight (kg)	477	3.34 (3.01, 3.69)	132	3.40 (3.06, 3.71)	3.00E-5	-0.87*	0.384*
Birth weight z-score	477	-0.16 (-0.88, 0.49)	132	-0.10 (-0.76, 0.46)	0.025	-0.53*	0.594*
Thrive index 12 months	450	0.15 (-0.49, 0.85)	124	0.14 (-0.39, 0.98)	0.983	-0.15†	0.878†
TACSTD2							
Gestational age (days)	519	40.00 (38.00, 41.00)	90	40.00 (39.00, 41.00)	1.04E-6	-1.49*	0.135*
Birth weight (kg)	519	3.35 (3.00, 3.69)	90	3.37 (3.12, 3.72)	3.00E-5	-1.29*	0.196*
Birth weight z-score	519	-0.16 (-0.90, 0.46)	90	-0.105 (-0.76, 0.55)	0.025	-1.00*	0.318*
Thrive index 12 months	490	0.15 (-0.47, 0.85)	84	0.11 (-0.60, 0.89)	0.983	0.90†	0.367†

Table 7.2 Comparison between early life exposures in whole GMS dataset and utilized dataset for both *IGF2* and *TACSTD2*.
‡ Shapiro-Wilk data normal test, †T test, *Mann Whitney U test

	Both genes		*IGF2* only		*TACSTD2* only		Comparison	
Variable	n	Median (IQR)	n	Median (IQR)	n	Median (IQR)	Stat (z or t)	p-value
Gestational age (days)	29	40.00 (39.00, 41.00)	103	40.00 (39.00, 41.00)	61	40.00 (39.00, 40.00)	0.10*	0.922*
Birth weight (kg)	29	3.37 (3.00, 3.67)	103	3.42 (3.06, 3.74)	61	3.37 (3.12, 3.77)	-0.56*	0.577*
Birth weight z-score	29	-0.12 (-0.76, 0.43)	103	-0.08 (-0.76, 0.50)	61	-0.09 (-0.74, 0.64)	-0.51*	0.613*
Thrive index 12 months	28	0.09 (-0.675, 1.05)	96	0.14 (-0.34, 0.95)	56	0.11 (-0.36, 0.83)	0.68†	0.497†

Table 7.3 Comparison between early life exposures in *IGF2* and *TACSTD2* datasets.
†T test, *Mann Whitney U test.

7.4.2 Early life characteristics in sample cohorts

A total of 132 individuals from the GMS cohort were assessed for *IGF2* DNA methylation while 90 individuals were assessed for *TACSTD2* DNA methylation. A total of 229 individuals from the NTFS cohort were measured for LINE-1 methylation. The baseline and early-life factors used in the analyses described are presented in Table 7.3. These variables were assessed for normality using the Shapiro-Wilk normal data test and a comparison of data between sexes was performed using a T test, Mann-Whitney U or Chi-squared test, as appropriate (Table 7.4).

The distribution of lifecourse exposure variables is shown in Table 7.3. Many of the variables showed non-normal distribution. Differences between male and female groups were observed for birth weight with a mean birthweight of 3.51kg (3.19, 3.75) in males and 3.29kg (2.98, 3.57) in females although this only attained statistical significance in the GMS *IGF2* subset (t=2.11, p=0.03).

	Males only		Females only		Normal test‡	Sex comparison	
Baseline variables at birth	n	Median (IQR)	n	Median (IQR)	p-value	Stat	p-value
GMS *IGF2* subset							
Gestational age (days)	68	40.00 (39.00, 41.00)	64	40.00 (39.00, 41.00)	0.001	0.01*	0.988*
Birth weight (kg)	68	3.51 (3.19, 3.75)	64	3.29 (2.98, 3.57)	0.003	2.11*	0.034*
Birth weight z-score	68	-0.06 (-0.62, 0.46)	64	-0.21 (-0.96, 0.46)	0.057	0.26†	0.795†
Thrive index 12 months	64	0.14 (-0.35, 0.95)	60	-0.07 (-0.43, 1.00)	0.108	0.30†	0.763†
Breastfeed (n) >4m/>6w/<6w/n	64	13/11/18/22	61	8/3/21/29	-	6.89$^{\chi}$	0.076$^{\chi}$
GMS *TACSTD2* subset							
Gestational age (days)	45	40.00 (39.00, 41.00)	45	40.00 (39.00, 41.00)	1.03E-4	0.31*	0.755*
Birth weight (kg)	45	3.51 (3.19, 3.78)	45	3.34 (2.88, 3.66)	0.044	1.66*	0.096*
Birth weight z-score	45	-0.07 (-0.74, 0.50)	45	-0.14 (-0.95, 0.55)	0.072	0.36†	0.721†
Thrive index 12 months	42	0.105 (-0.68, 0.89)	42	0.11 (-0.45, 0.89)	0.552	-0.56†	0.578†
Breastfeed (n) >4m/>6w/<6w/n	44	11/6/11/16	45	5/5/10/25	-	4.35$^{\chi}$	0.226$^{\chi}$
NTFS LINE-1 subset							
Gestational age (days)	83	40.00 (40.00, 40.00)	143	40.00 (40.00, 40.00)	1.02E-8	2.01*	0.044*
Birth weight (kg)	85	3.29 (3.03, 3.74)	144	3.36 (2.99, 3.74)	0.150	-0.38†	0.700†
Duration breastfed (weeks)	85	69.00 (28.00, 252.00)	139	60.00 (21.00, 165.00)	2.64E-6	1.17*	0.243*

Table 7.4 Lifecourse exposure variables measured in children at birth – *IGF2* and *TACSTD2* GMS subsets and NTFS LINE-1 subset.

Data shown for all of the individuals studied and stratified by sex. ‡ Shapiro-Wilk data normal test, †T test, *Mann Whitney U test, $^{\chi}$ Chi-squared.

7.4.3 Gene specific and genome wide DNA methylation

Gene specific methylation measures were utilized to investigate the relationship between early-life factors and DNA methylation at ages 8, and global LINE-1 DNA methylation assessed at age 49-51 years. Details of methods used to quantify DNA methylation are provided in Chapters 5 and 6. As previously discussed, it has been well documented that CpG sites within 1000bp of each other demonstrate strong correlation in methylation levels (Down et al., 2008). This was found to be the case in both *IGF2* and *TACSTD2* (Figure 7.4) in which three and seven sites respectively were highly positively correlated. As a result in the following analysis mean DNA methylation level was used to represent as all the measured CpG sites in each candidate gene.

DNA methylation was assessed in both gene specific assays by Pyrosequencing and these data are presented in Table 7.4. In the GMS cohort, *IGF2* showed an overall mean methylation of 42.72% (95% CI=38.72, 47.31), while *TACSTD2* showed an overall mean methylation of 41.10% (95% CI=29.48, 49.15) (Table 7.5).

LINE-1 global DNA methylation was also assessed in the NTFS cohort, with a mean overall methylation value of 52.74% (95% CI= 51.51, 54.89) observed (Table 7.5).

Sex differences were assessed by Mann-Whitney U test and no significant differences in DNA methylation were detectable between males and females in any CpG site in any of the amplicons assessed (Table 7.5).

	Boys only		Girls only		Normalcy test‡	Sex comparison	
Gene and CpG Site	n	Median Methylation (IQR) (%)	n	Median Methylation (IQR) (%)	p-value	Statistic (z or t)	p-value
IGF2							
Mean	68	42.01 (38.98, 46.19)	64	43.33 (38.49, 47.73)	4.20E-4	-0.75*	0.455*
TACSTD2							
Mean	45	39.83 (31.93, 47.38)	45	42.31 (29.43, 49.15)	0.001	0.35*	0.726*
LINE-1							
Mean	85	52.92 (51.55, 55.18)	144	52.53 (51.42, 54.80)	4.29E-10	0.65*	0.515*

Table 7.5 Mean *IGF2*, *TACSTD2* and LINE-1 DNA methylation in GMS children stratified by sex.

Data shown for all of the individuals studied and stratified by sex. ‡ Shapiro-Wilk data normal test, †T test, *Mann Whitney U test

7.4.4 Association analysis of DNA methylation against early life phenotypic variables

Association analyses were performed to assess the relationship between gene specific or global LINE-1 DNA methylation and early life exposures. Given the potential confounding effects of age and sex, multiple regression analyses were performed in order to adjust for these factors (Table 7.6). Significant associations were found between gestational age and DNA methylation in both *TACSTD2* in the GMS cohort (Coefficient= -1.77, p=0.04) and LINE-1 in the NTFS cohort (Coefficient= -2.05, p=0.04). For both associations, a shorter gestational age was related to an increase in methylation percentage. As indicated by the R^2 value, sex and gestational age accounted for ~5% and 2% of the variation in *TACSTD2* and LINE-1 methylation, respectively.

Gene and phenotypic trait	Coefficient [95% CI]	t	p-value	R^2 (%)
IGF2				
Gestational age (days)	-0.13 [-0.88, 0.61]	-0.35	0.729	0.01
Birth weight (kg)	0.57 [-1.76, 2.90]	0.49	0.627	0.01
Birth weight z-score	0.73 [-0.45, 1.90]	1.22	0.225	0.02
Thrive index 12 months	0.30 [-1.08, 1.67]	0.43	0.671	0.01
Breast feed > 4 months	Reference			
Breast feed > 6 weeks	1.63 [-3.37, 6.63]	0.64	0.520	
Breast feed < 6 weeks	0.37 [-3.56, 4.29]	0.18	0.854	0.01
Breast feed never	1.37 [-2.41, 5.14]	0.72	0.475	
TACSTD2				
Gestational age (days)	-1.77 [-3.44, -0.11]	-2.12	0.037	5.36
Birth weight (kg)	-2.53 [-7.12, 2.07]	-1.09	0.277	1.81
Birth weight z-score	-1.08 [-3.42, 1.25]	-0.92	0.358	1.43
Thrive index 12 months	0.45 [-2.36, 3.26]	0.32	0.750	0.82
Breast feed > 4 months	Reference			
Breast feed > 6 weeks	7.99 [-2.55, 18.54]	1.51	0.135	
Breast feed < 6 weeks	1.90 [-7.05, 10.86]	0.42	0.674	0.04
Breast feed never	4.59 [-3.51, 12.69]	1.13	0.263	
LINE-1				
Gestational age (days)	-3.00E-3 [-6.00E-3, -1.00E-4]	-2.05	0.042	2.14
Birth weight (kg)	0.07 [-0.29, 0.44]	0.40	0.689	0.42
Duration breastfed (weeks)	0.36 [-0.32, 1.04]	1.04	0.298	0.77

Table 7.6 Regression of early life phenotypic traits against global and gene specific DNA methylation.
Data shown adjusted for age and sex.

Given the non-normal distributions of methylation and phenotypic traits, further analysis was performed by Spearman's rank correlation stratifying for sex. Using this analysis a similar pattern of association was observed between *TACSTD2* DNA methylation and gestational age (all individuals Spearman's rho=-0.21, p-0.048), although this did not reach statistical significance in the smaller sex-stratified subgroups (Table 7.7). No other significant associations were observed between the early life exposures and *IGF2*, *TACSTD2* or LINE-1 DNA methylation (Table 7.6).

	All			Male			Female		
Gene and phenotypic trait	n	rho	p-value	n	rho	p-value	n	rho	p-value
IGF2									
Gestational age (days)	132	0.02	0.782	68	-0.02	0.882	64	0.06	0.653
Birth weight (kg)	132	0.05	0.550	68	-0.02	0.865	64	0.11	0.378
Birth weight z-score	132	0.11	0.224	68	0.07	0.554	64	0.12	0.330
Thrive index 12 months	124	-0.01	0.924	64	0.03	0.796	60	-0.02	0.908
Breast Feed‡	125	1.70	0.637	64	6.73	0.081	61	2.01	0.570
TACSTD2									
Gestational age (days)	90	-0.21	0.048	45	-0.21	0.161	45	-0.21	0.169
Birth weight (kg)	90	-0.09	0.406	45	-6.00E-3	0.965	45	-0.15	0.341
Birth weight z-score	90	-0.11	0.317	45	-0.07	0.671	45	-0.12	0.426
Thrive index 12 months	84	-0.04	0.739	42	0.15	0.342	42	-0.21	0.184
Breast Feed‡	89	3.85	0.278	44	1.81	0.614	45	3.67	0.300
LINE-1									
Gestational age (days)	226	0.03	0.667	83	0.01	0.921	143	0.03	0.740
Birth weight (kg)	229	-0.009	0.891	85	0.06	0.585	144	-0.05	0.550
Duration breastfed (weeks)	224	-0.10	0.138	85	-0.06	0.585	139	-0.14	0.110

Table 7.7 Spearman's rank correlation of early life phenotypic traits, gene specific and global DNA methylation.
Data shown stratified for sex. ‡ Kruskal–Wallis one-way analysis of variance

7.5 Discussion

These data demonstrate a correlation between gestational age at birth and gene specific (*TACSTD2*) DNA methylation in childhood and global (LINE-1) DNA methylation in later life. Since undertaking these analyses, additional data from our laboratory on an independent cohort has also demonstrated a link between gestational age and DNA methylation patterns measured at birth. Furthermore, Novakovic and colleagues recently reported widespread changes in promoter methylation profiles in human placenta in response to increasing gestational age (Novakovic et al., 2011). The observation that gestational age is associated with marked changes in epigenetic patterns aligns with existing knowledge in the field of developmental biology relating to the dynamic reprogramming of the epigenome during this period (Reik, 2007). What the lifecourse approach in the current study might suggest is that this dynamic *in utero* programming is somehow 'fixed' when the extrauterine environment is encountered and persists across the lifecourse. This postulate is somewhat speculative but provides an interesting paradigm for future work.

Global LINE-1 DNA methylation measured at birth is believed to be affected by several different exposures during pregnancy such as folate supplementation (Fryer et al., 2011) but also including tobacco smoke, polycyclic aromatic hydrocarbons (PAHs) (Herbstman et al., 2009) and perfluoroalkyl compounds (PFCs) (Guerrero-Preston et al., 2010). However, few studies to date have linked these or other exposures to methylation patterns later in the lifecourse, with the exception of the studies on the Dutch Hunger Winter cohort cited earlier (Heijmans et al., 2008, Tobi et al., 2009). In those studies linking early life exposure to later methylation levels, these have not yet uncovered a link to phenotypic traits that hints at the relationship not having a causal basis.

The data presented in this chapter could reflect the fact that differential DNA methylation is established, at least in part, *in utero* (Godfrey et al., 2011a), is fixed to some extent upon birth and persists through childhood and adulthood. Even if this is proven to be the case we are not yet able to conclude whether these persistent epigenetic changes are actually mechanistically related to disease risk.

However these observations are just postulates, further work will be required in order to replicate this observed association and delineate the causes of this differential methylation. This could be accomplished by performing gene specific and genome wide analyses on a variety of tissue types at a number of time points in large, well-powered studies.

Chapter 8: Discussion

8.1 Summary of aims and outcomes

A total of six aims were defined in this project. Key observations related to each of these aims are summarised and discussed below.

8.1.1 To identify tissue specific differential DNA methylation in rat hypothalamus in response to nutritional insult

The brain is a key area in the control of appetite and thus adiposity (Berthoud and Morrison, 2008). The arcuate nucleus of the hypothalamus integrates a variety of hormonal and autonomic signals to regulate feeding behaviour in mammals (Qi et al., 2010). A number of key appetite regulatory genes are expressed in arcuate nucleus neurones including *POMC* (Delahaye et al., 2008), NPY (Arai et al., 2010), AgRP (Briggs et al., 2010), and CART (Yoo et al., 2011) and changing expression levels of these genes may exert an effect on appetite regulation. DNA methylation levels are highly tissue specific and it has been postulated that methylation of gene promoters can play a role in the repression of gene transcription (Gibbs et al., 2010). Therefore DNA methylation was assessed in hypothalamic sections taken from a rat model of developmentally induced obesity. A primary antibody specific to 5-methylcytosine was selected alongside one specific to Beta III tubulin, a neurone specific cell cytoskeleton marker. These primary antibodies were labelled with fluorescent secondary antibodies to allow DNA methylation levels to be quantified. No significant differences in levels of 5-methylcytosine were found between the control group and either of the protein deplete or cross over (recuperated) groups. These findings suggest either that this methodology does not have enough sensitivity to detect differences in DNA methylation levels at a gross tissue level, or that no changes were present in response to the nutritional intervention. This finding would be considered contrary to recent studies that have shown that DNA methylation levels are altered in response to *in utero* dietary manipulation, albeit in gut as opposed to brain tissue, (McKay et al., 2011a) and that they can differ significantly in brain tissue in response to nutritional or hormonal interventions (Plagemann et al., 2009, Palou et al., 2011). This method results in the measurement of a cell specific methylation level with an average produced across an area of tissue. It is therefore effectively a global assessment of DNA methylation levels. This was a result of both the level of magnification available and the sensitivity of the fluorescent microscope. It may be that some level of mosaicism may operate whereby the

differences between DNA methylation levels vary between adjacent neurones in the same tissue and as such an assay assessing the whole area would mask any such changes. An assay that could assess DNA methylation in a neurone-by-neurone basis, perhaps utilising laser-capture micro-dissection, would allow this possibility to be explored.

8.1.2 To create a bioinformatic workflow allowing differentially methylated target genes to be identified from a gene expression array dataset

Target gene identification is a key step in any study into the epigenetic control of phenotype. Technology development has resulted in the generation of large datasets from expression studies that may provide a useful means of identifying candidate genes that may be regulated by DNA methylation levels. This study used a gene expression dataset that compared the offspring of mice fed either standard mouse chow or a highly palatable obesogenic diet. Differentially expressed genes were assessed using a number of bioinformatic tools in order to identify candidates that could potentially be regulated by DNA methylation. The gene list was mapped to biological pathways using the Ingenuity Pathway Analysis package. Those genes implicated in obesity related pathways remained in the analysis and were assessed for previous evidence of differential methylation using the MeInfoText tool. Candidates were then assessed for the presence of CpG islands within their promoters before pyrosequencing assay design was undertaken. Following these various bioinformatic workflow steps a total of four candidate genes were prioritised. These genes were *Esr1*, *Fxn*, *Igf2r*, and *Rbl2*. These four candidate genes were then assessed for DNA methylation levels using the pyrosequencing platform in the study reported in chapter 4. The attrition rate from this workflow was high and could impact on its application on smaller expression array datasets. However, changing the stringency level of a number of steps in the workflow including the gene having been previously implicated in differential methylation may result in a greater number of final candidates. The biological pathways polled in the pathway analysis step could also be changed to suit a differing study hypothesis. Another stage that could be modified is designing assays based around promoter sequences given recent data suggesting that inducers and exon one may be similarly implicated in the control of transcription. Finally, although CpG island DNA methylation has been investigated extensively, there is increasing interest in CpG island shores as sites of transcriptionally relevant DNA methylation (Dudziec et al., 2011, Cosgrove et al., 2011).

8.1.3 To quantify differential DNA methylation in target genes identified by the bioinformatic workflow

The identification of candidate differentially methylated genes from large expression array datasets can be a challenge, and the previous section sought to address this by applying a bioinformatic workflow. Candidates were identified that mapped to pathways implicated in the pathogenesis of obesity and related sequelae such as type 2 diabetes. This section of the study aimed to assess the methylation levels of the promoters of these identified target genes to discern if differential methylation was associated with the observed differential gene expression observed in the rodent model. Three of the four candidate genes analysed showed generally low levels of DNA methylation with *Esr1*, *Fxn* and *Rbl2* showing median methylation levels of approximately 5%. There was no significant difference between DNA methylation levels at these loci in control animals when compared to those exposed to a protein deplete diet. This finding does not rule out the possibility that promoter methylation acts as a transcriptional regulator for these genes, however it does show that there is a relatively low level of DNA methylation at the specific CpG sites assessed. In contrast, *Igf2r* however showed a median methylation level of approximately 50% as well as exhibiting significant differences between the offspring of control and nutritionally insulted animals. The level of methylation detected in the *Igf2r* promoter agrees with existing findings that *Igf2r* is maternally imprinted and therefore is hemimethylated (Stoger et al., 1993). The finding that *Igf2r* DNA methylation differed between experimental groups was interesting given the role of *Igf2r* in the control of cell growth and proliferation (Brown et al., 2009). Mice lacking *Igf2r* show foetal overgrowth compared with wild type mice (Wylie et al., 2003) and it is possible that differential methylation in the *Igf2r* promoter may regulate *Igf2r* expression. If this were the case increased DNA methylation could lead to repression of gene expression and the development of an obese phenotype. Further work is needed to link changes in *Igf2r* DNA methylation causally to the development of obesity, as it is possible that the change in methylation levels is mediated by the development of an obese phenotype (reverse causation). To assess this DNA methylation measurements could be taken prior to the development of an obese phenotype in a prospective model. The investigation of *Igf2r* methylation in this methodology was limited in scope by the amplicon length attainable using the pyrosequencing assay. Further studies could utilise techniques with increased read lengths such as the Sequenom platform to identify if the change in methylation observed in the section of the promoter assayed in this study maps to other transcriptionally relevant areas of the genome such as enhancer and first exon sequences.

8.1.4 To assess gene specific DNA methylation in relation to markers of phenotypic health at age eight years

There is a large body of evidence suggesting that DNA methylation can affect gene expression. However there is limited evidence outside of neoplasias linking DNA methylation to a phenotypic change. This section set out to assess if methylation of candidate genes that had previously been implicated in the pathogenesis of childhood obesity were related to markers of metabolic health in children. The promoter regions of the genes *IGF2* and *TACSTD2* were assessed for methylation using pyrosequencing of bisulfite modified DNA. Methylation levels were regressed against phenotypic traits including weight, height, bioimpedance and BMI. In the case of *TACSTD2* no significant associations were found between DNA methylation and phenotypic traits at age eight years. This is contrary to recent literature suggesting that *TACSTD2* methylation is associated with the risk of fat mass in children aged 11 (Groom et al., 2012), although this paper did acknowledge that this association was likely to be confounded and non-causal. The findings presented here might indicate that *TACSTD2* does not play a causal role in the development of obesity. However, *IGF2* methylation was significantly associated with the age in the individuals studied. This concurs with previous studies that have reported age related changes in DNA methylation at the *IGF2* locus (Maegawa et al., 2010, Heijmans et al., 2008), even if in the case of this study the age range studied is relatively small. The level of plasticity and overall trajectory of this methylation level as the lifecourse progresses is however difficult to predict. Current theories suggest that the variability of methylation at this locus could be expected to increase with age due to aberrant maintenance of methylation by DNMTs gradually decreasing in efficiency (Xiao et al., 2008). The association between *IGF2* promoter methylation and age in this study may be attributable to environmental and stochastic factors. The lack of association with potential markers of metabolic disease in *IGF2* however is contrary to recent reports linking *IGF2* methylation with the development of obesity. Further studies in this area might consider a greater range of markers of metabolic disease in addition to those used in this study, as well as a wider range of phenotype and age in order to reveal a relationship between DNA methylation and obesity. Given the role of *IGF2* in insulin signalling it would be interesting to assess whether there is a link between *IGF2* DNA methylation and markers of glucose tolerance such fasting insulin and glucose concentrations and HOMA-IR.

8.1.5 To assess the effect of global DNA methylation at age 50 years on markers of metabolic health

There is increasing evidence linking levels of global DNA methylation with phenotypic traits that could be indicative of early stage disease (Rabinovich et al., 2010). A number of assays can be used to assess global DNA methylation, one of which is a PCR-based assay specific to Long interspersed nuclear element 1 (LINE-1). LINE-1 is a repetitive, or transposable, element present throughout the genome that has been implicated in the control of cell division (Singer et al., 2010). This section of the study aimed to assess whether global DNA methylation measured at age 50 years is associated with traits indicative of early stage metabolic disease. To this end single base resolution sequencing of LINE-1 methylation was undertaken. These measurements were then assessed with respect to cardiometabolic traits. Following multiple regression analyses, with adjustment for sex, increased levels of LINE-1 DNA methylation were associated with increased fasting glucose, total cholesterol, total triglycerides and LDL cholesterol, and with decreasing HDL cholesterol and HDL:LDL ratio. These metabolic markers may be indicative of both a change in the glucose/insulin signalling pathway and early stage dyslipidaemia. Therefore when taken together, these changes are suggestive of a pre-metabolic disease phenotype. This finding is interesting given evidence suggesting that global DNA methylation is associated with both lipid profiles (Cash et al., 2011) and cardiovascular disease risk (Kim et al., 2010). However, it is important to remember that LINE-1 methylation is only a surrogate for global DNA methylation. It would be interesting to assess other repetitive elements such at the *Alu* repeat and the Short interspersed nuclear elements (SINEs) to see if this observation could be replicated. Given the metabolic traits with which LINE-1 DNA methylation appears to be associated, gene specific assays focussing on the glucose and lipid homeostasis pathways might show that changes in the promoter regions of functional genes have an effect on metabolic disease risk. It would also be worth investigating whether any of the individuals in this study shown to be exhibiting early stage metabolic disease phenotypes progress to develop type 2 diabetes and related sequelae. Indeed, the Newcastle Thousand Families Study cohort is being followed up currently at age 60 years with both clinical and biological measurements and the prognostic value of theLINE-1 measures taken at age 50 can be evaluated. Further, it would allow the temporal changes in global DNA methylation to be tracked to metabolic change, demonstrating a potential causal link between differential DNA methylation and the development of metabolic disease.

8.1.6 To assess the effect of early life exposures on DNA methylation levels in later life

Studies have shown that early life exposures can exert an effect on DNA methylation levels. Nutritional (McKay et al., 2011b) and environmental exposure such as PAHs (Herbstman et al., 2012) and PFOS (Guerrero-Preston et al., 2010) during the critical windows *in utero* and early postnatal life have been linked with differential DNA methylation. Less is known however about the persistence of these epigenetic changes throughout the life course. This section of the study set out to test the hypothesis that early life influences can affect DNA methylation levels and that these epigenetic changes persist in later life. Early growth and nutrition were assessed in both the Thousand Families Study and Gateshead Millennium Study cohort through measures such as gestational age, birth weight and duration of breastfeeding. These measures were then regressed against a number of DNA methylation measurements including global methylation, as measured by the LINE-1 assay, and the gene-specific assays *IGF2* and *TACSTD2*. These measures were of particular interest given recent literature showing differential methylation at the *IGF2* locus in children born small for gestational age (Tobi et al., 2011) and significantly lower LINE-1 global methylation levels in new born babies with low or high birth weight when compared with normal weight infants (Michels et al., 2011). Gene specific *TACSTD2* DNA methylation levels were significantly associated with gestational age in the GMS cohort, whereas LINE-1 global DNA methylation was shown to be significantly associated with gestational age in the NTFS cohort. In both cases, a shorter gestational age was associated with increased DNA methylation. These results are particularly interesting given recent reports suggesting that promoter DNA methylation profiles alter in human placenta in response to increasing gestational age (Novakovic et al., 2011). Taken together the findings of this section of the study suggest that differential DNA methylation may be established *in utero* and persist through childhood and adulthood. Further work will be required to elucidate whether these persistent epigenetic changes are linked causally to increased risk of developing obesity and related comorbidities.

8.2 Key questions for future work

The role of epigenetic mechanisms in the developmental programming of obesity and related co-morbidities has been explored in this thesis. The discipline of epigenetics and its application to understanding the aetiology of common complex diseases such as obesity is a nascent area of investigation. A number of important questions remain to be addressed as this field of research evolves. The work conducted in my thesis serves to highlight several of these questions which are summarised below.

8.2.1 When are the 'critical windows' in development in terms of epigenetic vulnerability?

The role of early life exposures in the developmental programming of many conditions including obesity and related metabolic diseases is well characterised. The work presented in my thesis postulates that DNA methylation is a key step in the mediation of this process and posits that *in utero* and early postnatal influences may be of greater importance than exposures encountered at other points in the life course. However, these may not necessarily be the only, or even the most important, determinants of epigenetic patterns. There is evidence that the pubertal period may be an example of a 'critical window' in development and that changes established at this point may persist throughout the lifecourse (Jasik and Lustig, 2008). Pearce et al (2012) also present evidence, based on observations made in the Newcastle Thousand Families Study that influences in adulthood are probably more important determinants of later disease risk that those experienced in early life (Pearce et al., 2012). Whether these 'critical windows' are the same as those for epigenetic programming remains to be determined. Assessing the role of epigenetic variation at multiple points throughout the life course and not focusing exclusively on the early life period may help to elucidate more fully the role of DNA methylation in the developmental programming of obesity. Although there is some evidence in animal models that i*n utero* exposure to agents such as the methyl donor genistein can program patterns of DNA methylation in later life (Dolinoy et al., 2006), the stability of these patterns throughout the life course remains less well characterised. Further work in this area should address whether DNA methylation changes programmed during the critical window of *in utero* and early post natal life persist throughout adulthood and exert an effect on gene expression levels and therefore the phenotype.

8.2.2 Which tissue type should we be assessing?

DNA methylation and gene expression patterns are tissue specific, indeed epigenetic processes are the mechanism driving tissue differentiation (Doi et al., 2009). The issue therefore arises that if a particular disease or disorder is known to act through perturbation of a specific cell or tissue type, how informative is it to study epigenetic patterns in a surrogate non-target tissue?

The health outcome of interest in this study was obesity and its related comorbidities including type 2 diabetes and cardiovascular disease. This should inform the target tissues of interest, and in the case of the analyses performed on animal models, it did. Heart, liver, white adipose tissue and brain were selected for various analyses of gene expression and DNA methylation and this makes sense given the tissue-specific phenotypic traits of cardiovascular disease, NAFLD, perturbed adiposity and altered appetite regulation respectively. This issue becomes more complex however when we consider the human studies. For both practical and ethical reasons, human DNA samples are generally taken as non-invasively as possible. Saliva, buccal swabs, blood samples or, where there is an underlying pathology, biopsy are the usual sources of DNA. It is very difficult to obtain samples of other human tissues for obvious ethical reasons. As such there is an issue of comparability between studies in animals (or humans) using specific tissue types and in those using blood or saliva (McKay et al., 2011c). Recent studies have indicated that peripheral blood may be a suitable surrogate tissue in the context of obesity (Relton et al., 2012), where those genes identified as being differentially methylated showed some overlap with genes identified in other published studies of gene expression in adipose tissue (Pietiläinen et al., 2006). DNA from peripheral blood and saliva have been utilised in the work presented in this thesis, with limitations recognised regarding the inferences that can be made. As the field advances more work needs to be done to ascertain if DNA extracted from saliva and blood is a good epigenetic surrogate for a specific obesity-relevant target tissue such as adipose, muscle, liver and brain.

8.2.3 Is cell heterogeneity an issue?

Even tissues that are relatively homogenous contain a range of different cell types, for instance buccal swabs contain both buccal epithelial cells and white blood cells. When a sample has been extracted from a complex multicellular tissue such as brain or liver the variety of cell types involved increases markedly. This is key because each cell type may have a subtly different epigenetic signature and gene expression profile depending on its

function. Performing methylation analysis on a DNA sample extracted from a multicellular tissue sample will result in an average DNA methylation profile for that tissue. While this measure may be of interest it is possible that high and low levels of methylation in distinct cell types combined with changes in the proportions of these which are present will result in masking of any differences in DNA methylation present due to the factor of interest. Recent studies have addressed this question by assessing DNA methylation levels across a range of blood cell types including granulocytes and mononucleocytes (Wu et al., 2011b). Approaches such as accounting for differential blood cell count and ensuring consistency in the location and composition of any biopsy sample will help in controlling for this inherent source of variation.

8.2.4 How should we go about selecting candidate genes?

Candidate gene selection has been a key component of the study designs presented in my thesis. High throughput systems have often been used to produce gene lists from which the candidates are selected, and picking candidates from such lists can be challenging. The process relies upon existing, sometimes imperfect, information, precluding the prioritisation of potentially novel loci. It is also vulnerable to subjectivity and bias. In this thesis I have presented a bioinformatic workflow that attempts to incorporate the steps required to identify candidate genes relevant to the pathogenesis of obesity. However, despite utilising state-of-the-art bioinformatics tools, this process is not without limitations. As high throughput analysis continues to scale up and generate an increasing archive of multi-dimensional data, so the size of the gene lists of potential target genes will grow. Future work on DNA methylation will require more extensive bioinformatic analysis of both gene expression array datasets and high throughput methylation datasets to identify pathways and genes of interest.

8.2.5 Where should we be looking in the gene?

The work presented in my thesis has focussed on promoter regions of candidate genes when assessing DNA methylation levels, due to their well-characterised role in the regulation of gene transcription. This is an approach that has been utilised widely in the study of functionally relevant DNA methylation. However it is important to note that there are a number of alternative regions of the gene that play a role in transcriptional regulation. These regions may also be candidates for differential DNA methylation. Regions such as silencers, insulators and enhancers are known to play a role in transcriptional regulation (Lupien et al., 2008). There is increasing evidence to suggest that differential DNA

methylation is involved in this process. For example one study has shown that decreased enhancer DNA methylation is associated with increased gene expression (Xu et al., 2007). The promoter is still a key area of interest but future studies into the role of DNA methylation in the control of transcription will undoubtedly extend beyond promoter regions to investigate some of these alternative genomic regions.

8.2.6 How relevant are CpG islands?

For many years CpG islands have been the focus of studies into DNA methylation levels due to the high density of DNA methylation marks. However in recent years the study of areas of the gene some two kb distant from CpG islands, dubbed 'CpG island shores', has shown that these areas are more likely to display variation in DNA methylation than CpG islands themselves. Moreover, there is evidence that CpG islands falling within promoter sequences are not the only ones of functional relevance. CpG islands located either between genes or within a transcribed section of DNA have been shown to exhibit a high degree of tissue-specific methylation (Illingworth et al. 2008; Rauch et al. 2009; Maunakea et al. 2010). Known as 'Orphan' CGIs, these account for some 50% of the CpG islands in both human and mouse genomes (Illingworth et al. 2010). Though Orphan CGIs do not contain recognised promoter sequences they have been shown to recruit RNA polymerase II resulting in transcribed, possibly non-coding RNA (Illingworth et al., 2010). Therefore an interesting future direction of study could be to assess the levels of DNA methylation in both CpG island shores and Orphan CGIs in order to discern any possible role in transcriptional regulation and ultimately functional consequences for disease risk.

8.2.7 Does differential methylation cause differential expression?

For an epigenetic epidemiological study to be able to determine that DNA methylation is causally linked to a change it has to be demonstrated that it is present concomitantly with a change in gene expression levels. This was the philosophy behind Chapters 3 and 4 of my thesis, where the promoter methylation levels of differentially expressed genes were assessed. However the ideal experimental design would involve overlaying gene expression data with DNA methylation data from the same sample source (McKay et al., 2008). This would allow changes in DNA methylation level to be causally linked to changes in gene expression. More elaborate functional studies could be applied as an adjunct to epidemiological approaches to define causal relationships. These might include reporter gene assays or cell culture experiments utilising demethylating agents to establish the relationship between DNA methylation and gene expression (Hitchins et al., 2011).

8.2.8 How important are other epigenetic modifications?

DNA methylation is only one of a range of epigenetic modifications that have been shown to exert an effect on gene expression levels and possibly phenotype. Histone modifications and ncRNAs have also been shown to be involved in transcriptional regulation. There has also been a recent upsurge of interest in the field of RNA epigenetics that is post-transcriptional modification of the RNA (He, 2010). These modifications include RNA methylation and it has been suggested that they may have functions beyond structural maintenance of the RNA. Understanding the role of these other epigenetic modifications alongside DNA methylation would help us to more fully unravel the role of developmental programming in the pathogenesis of obesity.

8.3 Conclusions

In summary, the work presented in my thesis aimed to further understand the role of DNA methylation as a mechanism in the developmental programming of obesity by utilising both animal models and human cohorts. These two approaches were used in a complementary way, with data generated in animal models used to inform candidate gene studies in humans. The work incorporated a range of techniques including, bioinformatics, immunohistochemistry, quantitative DNA methylation analysis and statistical analysis methods pertinent to these approaches.

Although the work presented in this thesis was not able to confirm diet-induced differences in DNA methylation in the arcurate nuclei of rodents, the hypothalamic axis undoubtedly plays a key role in the moderation of appetite and therefore adiposity and further exploration of epigenetic signatures in this region are warranted. The absence of an association between the groups analysed here may be due to the sensitivity of the technique to detect differences and cannot be interpreted as the absence of any epigenetic phenomena.

A novel bioinformatic workflow to identify differentially methylated genes from a gene expression array dataset was presented and four candidates selected. The *Igf2r* gene identified in rodents using a bioinformatic workflow was shown to exhibit differential DNA methylation in liver tissue of offspring exposed *in utero* to a maternal obesogenic diet.

In humans, increased *IGF2* promoter DNA methylation was found to be positively associated with age in a cohort of children. In a cohort of individuals aged 50 years,

increased global LINE-1 DNA methylation was found to be associated with markers of cardiometabolic risk in later life, revealing the possibility that this may be a useful. biomarker of future disease risk. Increased levels of DNA methylation in the *TACSTD2* and *IGF2* promoters were also found to be associated with gestational age in cohorts aged eight and 50 respectively. This study has shown that differential DNA methylation is associated with potential biomarkers of metabolic disease in humans. Additional research in this area should help further elucidate the role of DNA methylation in the developmental programming of obesity.

References

Aguilera, C., Gil-Campos, M. and Canete, R. (2008) 'Alterations in plasma and tissue lipids associated with obesity and metabolic syndrome', *Clinical Science,* 114, pp. 183-193.

Al-Mahdawi, S., Pinto, R. M., Ismail, O., Varshney, D., Lymperi, S., Sandi, C., Trabzuni, D. and Pook, M. (2008) 'The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues', *Human molecular genetics,* 17, (5), pp. 735.

Ali, S. and Coombes, R. C. (2000) 'Estrogen receptor alpha in human breast cancer: occurrence and significance', *Journal of mammary gland biology and neoplasia,* 5, (3), pp. 271-281.

Allender, S. and Rayner, M. (2011) 'Assessing the burden of overweight and obesity: an example from the United Kingdom', *Obesity facts,* 2, (2), pp. 8-8.

Almen, M. S., Jacobsson, J. A., Moschonis, G., Benedict, C., Chrousos, G. P., Fredriksson, R. and Schroth, H. B. (2012) 'Genome Wide Analysis Reveals Association of a FTO Gene Variant with Epigenetic Changes', *Genomics*, 99, (3), pp. 132-7.

An, J. Y., Kim, E. A., Jiang, Y., Zakrzewska, A., Kim, D. E., Lee, M. J., Mook-Jung, I., Zhang, Y. and Kwon, Y. T. (2010) 'UBR2 mediates transcriptional silencing during spermatogenesis via histone ubiquitination', *Proceedings of the National Academy of Sciences,* 107, (5), pp. 1912-1917.

An, W. (2007) 'Histone acetylation and methylation: combinatorial players for transcriptional regulation', *Sub-Cellular Biochemistry,* 41, pp. 351-69.

Angulo, P. (2007) 'Obesity and nonalcoholic fatty liver disease', *Nutrition reviews,* 65, pp. S57-S63.

Arai, Y., Gradwohl, G. and Kameda, Y. (2010) 'Expression of neuropeptide Y and agouti-related peptide in the hypothalamic arcuate nucleus of newborn neurogenin3 null mutant mice', *Cell and Tissue Research,* 340, (1), pp. 137-145.

Arenz, S., Rockerl, R., Koletzko, B. and Von Kries, R. (2004) 'Breast-feeding and childhood obesity - a systematic review', *International Journal of Obesity,* 28, (10), pp. 1247-1256.

Arnaud, P., Goubely, C., Pelissier, T. and Deragon, J. M. (2000) 'SINE retroposons can be used in vivo as nucleation centers for de novo methylation', *Molecular and cellular biology,* 20, (10), pp. 3434.

Aubert, R., Suquet, J. P. and Lemonnier, D. (1980) 'Long-term morphological and metabolic effects of early under-and over-nutrition in mice', *The Journal of nutrition,* 110, (4), pp. 649.

Baccarelli, A., Tarantini, L., Wright, R. O., Bollati, V., Litonjua, A. A., Zanobetti, A., Sparrow, D., Vokonas, P. and Schwartz, J. (2010a) 'Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study', *Epigenetics: Official Journal of the DNA Methylation Society,* 5, (3).

Baccarelli, A., Wright, R., Bollati, V., Litonjua, A., Zanobetti, A., Tarantini, L., Sparrow, D., Vokonas, P. and Schwartz, J. (2010b) 'Ischemic heart disease and stroke in relation to blood DNA methylation', *Epidemiology,* 21, (6), pp. 819-28.

Baccarelli, A., Wright, R. O., Bollati, V., Tarantini, L., Litonjua, A. A., Suh, H. H., Zanobetti, A., Sparrow, D., Vokonas, P. S. and Schwartz, J. (2009) 'Rapid DNA methylation changes after exposure to traffic particles', *Am J Respir Crit Care Med,* 179, (19136372), pp. 572-578.

Baghi, R. (2009) 'DNA methylation effects in Friedreich ataxia (FRDA)'.

Bannister, A. J. and Kouzarides, T. (2011) 'Regulation of chromatin by histone modifications', *Cell research,* 21, (3), pp. 381-395.

Barker, D. J. P., Godfrey, K. M., Gluckman, P., Harding, J. E., Owens, J. A. and Robinson, J. S. (1993) 'Fetal nutrition and cardiovascular disease in adult life', *The Lancet,* 341, (8850), pp. 938-941.

Barter, P., Gotto, A. M., LaRosa, J. C., Maroni, J., Szarek, M., Grundy, S. M., Kastelein, J. J. P., Bittner, V. and Fruchart, J. C. (2007) 'HDL cholesterol, very low levels of LDL cholesterol, and cardiovascular events', *New England Journal of Medicine,* 357, (13), pp. 1301-1310.

Barth, T. K. and Imhof, A. (2010) 'Fast signals and slow marks: the dynamics of histone modifications', *Trends in biochemical sciences,* 35, (11), pp. 618-626.

Bauer, U. M., Daujat, S., Nielsen, S. J., Nightingale, K. and Kouzarides, T. (2002) 'Methylation at arginine 17 of histone H3 is linked to gene activation', *EMBO reports,* 3, (1), pp. 39-44.

Bayol, S., Simbi, B., Bertrand, J. and Stickland, N. (2008) 'Offspring from mothers fed a 'junk food' diet in pregnancy and lactation exhibit exacerbated adiposity that is more pronounced in females', *The Journal of Physiology,* 586, (13), pp. 3219-3230.

Begum, G., Stevens, A., Smith, E. B., Connor, K., Challis, J. R. G., Bloomfield, F. and White, A. (2012) 'Epigenetic changes in fetal hypothalamic energy regulating pathways are associated with maternal undernutrition and twinning', *The FASEB Journal*, 26, (4), pp. 1694-703.

Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R. and Deininger, P. (2010) 'Somatic expression of LINE-1 elements in human tissues', *Nucleic Acids Res,* 38, (20215437), pp. 3909-3922.

Belgnaoui, S. M., Gosden, R. G., Semmes, O. J. and Haoudi, A. (2006) 'Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells', *Cancer Cell Int,* 6, (16670018), pp. 13-13.

Bell, A. C. and Felsenfeld, G. (2000) 'Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene', *Nature,* 405, (6785), pp. 482-485.

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., Gilad, Y. and Pritchard, J. K. (2011) 'DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines', *Genome Biology,* 12, (1), pp. R10.

Bellinger, L., Lilley, C. and 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', *British Journal of Nutrition,* 92, pp. 513-520.

Ben-Shlomo, Y. and Kuh, D. (2002) 'A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives', *International Journal of Epidemiology,* 31, (2), pp. 285.

Benetti, R., Gonzalo, S., Jaco, I., MuÒoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T. and Klatt, P. (2008) 'A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases', *Nature structural & molecular biology,* 15, (3), pp. 268.

Berenson, G. S. (2012) 'Health Consequences of Obesity', *Pediatric Blood & Cancer,* 58, (1), pp. 117-121.

Berger, S. L., Kouzarides, T., Shiekhattar, R. and Shilatifard, A. (2009) 'An operational definition of epigenetics', *Genes & development,* 23, (7), pp. 781.

Bergink, S. and Jentsch, S. (2009) 'Principles of ubiquitin and SUMO modifications in DNA repair', *Nature,* 458, (7237), pp. 461-467.

Berisha, S. Z., Serre, D., Schauer, P., Kashyap, S. R. and Smith, J. D. (2011) 'Changes in Whole Blood Gene Expression in Obese Subjects with Type 2 Diabetes Following Bariatric Surgery: a Pilot Study', *PloS one,* 6, (3), pp. e16729.

Bermudez-Silva, F. J., Cardinal, P. and Cota, D. (2012) 'The role of the endocannabinoid system in the neuroendocrine regulation of energy balance', *Journal of Psychopharmacology,* 26, (1), pp. 114-124.

Bernstein, B. E., Meissner, A. and Lander, E. S. (2007) 'The mammalian epigenome', *Cell,* 128, (4), pp. 669-681.

Berridge, K. C., Ho, C. Y., Richard, J. M. and DiFeliceantonio, A. G. (2010) 'The tempted brain eats: pleasure and desire circuits in obesity and eating disorders', *Brain research,* 1350, pp. 43-64.

Berthoud, H. R. (2004) 'Neural control of appetite: cross-talk between homeostatic and non-homeostatic systems', *Appetite,* 43, (3), pp. 315-317.

Berthoud, H. R. and Morrison, C. (2008) 'The brain, appetite, and obesity', *Annu. Rev. Psychol.,* 59, pp. 55-92.

Bird, A. (2002) 'DNA methylation patterns and epigenetic memory', *Genes & development,* 16, (1), pp. 6.

Bird, A. P. and Wolffe, A. P. (1999) 'Methylation-induced repression--belts, braces, and chromatin', *Cell,* 99, (5), pp. 451.

Bjornsson, H. T., Sigurdsson, M. I., Fallin, M. D., Irizarry, R. A., Aspelund, T., Cui, H., Yu, W., Rongione, M. A., Ekstr√∂m, T. J. and Harris, T. B. (2008) 'Intra-individual change over time in DNA methylation with familial clustering', *JAMA: the journal of the American Medical Association,* 299, (24), pp. 2877.

Bjornstrom, L. and Sjoberg, M. (2005) 'Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes', *Molecular Endocrinology,* 19, (4), pp. 833-842.

Black, A. E. (2000) 'Critical evaluation of energy intake using the Goldberg cut-off for energy intake: basal metabolic rate. A practical guide to its calculation, use and limitations', *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity,* 24, (9), pp. 1119.

Blevins, J. and Baskin, D. (2010) 'Hypothalamic-brainstem circuits controlling eating'.

Blondeau, B., Avril, I., Duchene, B. and Breant, B. (2002) 'Endocrine pancreas development is altered in foetuses from rats previously showing intra-uterine growth retardation in response to malnutrition', *Diabetologia,* 45, (3), pp. 394-401.

Bo, S., Cavallo-Perin, P., Scaglione, L., Ciccone, G. and Pagano, G. (2000) 'Low birthweight and metabolic abnormalities in twins with increased susceptibility to Type 2 diabetes mellitus', *Diabetic medicine,* 17, (5), pp. 365-370.

Boden, G. (2011) 'Obesity, insulin resistance and free fatty acids', *Current Opinion in Endocrinology, Diabetes and Obesity,* 18, (2), pp. 139.

Bollati, V., Galimberti, D., Pergoli, L., Dalla Valle, E., Barretta, F., Cortini, F., Scarpini, E., Bertazzi, P. A. and Baccarelli, A. (2011) 'DNA methylation in repetitive elements and Alzheimer disease', *Brain, Behavior, & Immunity,* 25, (6), pp. 1078-83.

Boney, C. M., Verma, A., Tucker, R. and Vohr, B. R. (2005) 'Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus', *Pediatrics,* 115, (3), pp. e290.

Bosserhoff, A. and Hellerbrand, C. (2011) 'Obesity and Fatty Liver Are 'Grease' for the Machinery of Hepatic Fibrosis', *Digestive Diseases,* 29, (4), pp. 377-383.

Boston, B. A. (2001) 'Pro-opiomelanocortin and weight regulation: from mice to men', *Journal of pediatric endocrinology & metabolism: JPEM,* 14, pp. 1409.

Bouret, S. G. (2009) 'Early life origins of obesity: role of hypothalamic programming', *Journal of pediatric gastroenterology and nutrition,* 48, pp. S31.

Bouret, S. G., Gorski, J. N., Patterson, C. M., Chen, S., Levin, B. E. and Simerly, R. B. (2008) 'Hypothalamic neural projections are permanently disrupted in diet-induced obese rats', *Cell Metabolism,* 7, (2), pp. 179-185.

Boyes, J. and Bird, A. (1991) 'DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein', *Cell,* 64, (6), pp. 1123-1134.

Branciamore, S., Chen, Z. X., Riggs, A. D. and Rodin, S. N. (2010) 'CpG island clusters and pro-epigenetic selection for CpGs in protein-coding exons of HOX and other transcription factors', *Proceedings of the National Academy of Sciences,* 107, (35), pp. 15485-15490.

Braulke, T. (1999) 'Type-2 IGF receptor: a multi-ligand binding protein', *Hormone and metabolic research,* 31, (2), pp. 242.

Breslow, N. E. and Day, N. E. (1987) 'Statistical methods in cancer research. Volume II--The design and analysis of cohort studies', *IARC scientific publications*, (82), pp. 1.

Breton, C. V., Byun, H.-M., Wenten, M., Pan, F., Yang, A. and Gilliland, F. D. (2009) 'Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation', *Am J Respir Crit Care Med,* 180, (19498054), pp. 462-467.

Briggs, D. I., Enriori, P. J., Lemus, M. B., Cowley, M. A. and Andrews, Z. B. (2010) 'Diet-induced obesity causes ghrelin resistance in arcuate NPY/AgRP neurons', *Endocrinology,* 151, (10), pp. 4745.

Brion, M. J. A., Lawlor, D. A., Matijasevich, A., Horta, B., Anselmi, L., Ara√ʃjo, C. L., Menezes, A. M. B., Victora, C. G. and Smith, G. D. (2011) 'What are the causal effects of breastfeeding on IQ, obesity and blood pressure? Evidence from comparing high-income with middle-income cohorts', *International journal of epidemiology,* 40, (3), pp. 670-680.

Brooks, J. D., Cairns, P., Shore, R. E., Klein, C. B., Wirgin, I., Afanasyeva, Y. and Zeleniuch-Jacquotte, A. (2010) 'DNA methylation in pre-diagnostic serum samples of breast cancer cases: Results of a nested case-control study', *Cancer epidemiology,* 34, (6), pp. 717-723.

Brown, J., Jones, E. Y. and Forbes, B. E. (2009) 'Keeping IGF-II under control: Lessons from the IGF-II-IGF2R crystal structure', *Trends in biochemical sciences,* 34, (12), pp. 612-619.

Brown, S. E., Weaver, I. C. G., Meaney, M. J. and Szyf, M. (2008) 'Regional-specific global cytosine methylation and DNA methyltransferase expression in the adult rat hippocampus', *Neuroscience letters,* 440, (1), pp. 49-53.

Buchwald, H., Avidor, Y., Braunwald, E., Jensen, M. D., Pories, W., Fahrbach, K. and Schoelles, K. (2004) 'Bariatric surgery', *JAMA: The Journal of the American Medical Association,* 292, (14), pp. 1724.

Buettner, C., Muse, E. D., Cheng, A., Chen, L., Scherer, T., Pocai, A., Su, K., Cheng, B., Li, X. and Harvey-White, J. (2008) 'Leptin controls adipose tissue lipogenesis via central, STAT3 - independent mechanisms', *Nature medicine,* 14, (6), pp. 667-675.

Bulteau, A. L., O'Neill, H. A., Kennedy, M. C., Ikeda-Saito, M., Isaya, G. and Szweda, L. I. (2004) 'Frataxin acts as an iron chaperone protein to modulate mitochondrial aconitase activity', *Science,* 305, (5681), pp. 242.

Burdge, G. C., Tricon, S., Morgan, R., Kliem, K. E., Childs, C., Jones, E., Russell, J. J., Grimble, R. F., Williams, C. M. and Yaqoob, P. (2005) 'Incorporation of cis-9, trans-11 conjugated linoleic acid and vaccenic acid (trans-11 18: 1) into plasma and leukocyte lipids in healthy men consuming dairy products naturally enriched in these fatty acids', *British Journal of Nutrition,* 94, (2), pp. 237-243.

Byun, H. M., Siegmund, K. D., Pan, F., Weisenberger, D. J., Kanel, G., Laird, P. W. and Yang, A. S. (2009) 'Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue-and individual-specific DNA methylation patterns', *Human molecular genetics,* 18, (24), pp. 4808.

Cali, A. M. G. and Caprio, S. (2009) 'Ectopic Fat Deposition, Adiponectin and Insulin Resistance in Obese Adolescents', *Insulin Resistance*, pp. 149-159.

Calvanese, V., Lara, E., Kahn, A. and Fraga, M. F. (2009) 'The role of epigenetics in aging and age-related diseases', *Ageing research reviews,* 8, (4), pp. 268-276.

Calvano, S. E., Xiao, W., Richards, D. R., Felciano, R. M., Baker, H. V., Cho, R. J., Chen, R. O., Brownstein, B. H., Cobb, J. P. and Tschoeke, S. K. (2005) 'A network-based analysis of systemic inflammation in humans', *Nature,* 437, (7061), pp. 1032.

Campos, E. I. and Reinberg, D. (2009) 'Histones: annotating chromatin', *Annual review of genetics,* 43, pp. 559-599.

Cao, J., Zhou, J., Gao, Y., Gu, L., Meng, H., Liu, H. and Deng, D. (2009) 'Methylation of p16 CpG island associated with malignant progression of oral epithelial dysplasia: a prospective cohort study', *Clinical Cancer Research,* 15, (16), pp. 5178-5183.

Cash, H. L., McGarvey, S. T., Houseman, E. A., Marsit, C. J., Hawley, N. L., Lambert-Messerlian, G. M., Viali, S., Tuitele, J. and Kelsey, K. T. (2011) 'Cardiovascular disease risk factors and DNA methylation at the LINE-1 repeat region in peripheral blood from Samoan Islanders', *Epigenetics: official journal of the DNA Methylation Society,* 6, (10).

Castaldo, I., Pinelli, M., Monticelli, A., Acquaviva, F., Giacchetti, M., Filla, A., Sacchetti, S., Keller, S., Avvedimento, V. E. and Chiariotti, L. (2008) 'DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients', *Journal of medical genetics,* 45, (12), pp. 808.

Catalan, V., Gomez-Ambrosi, J., Rodriguez, A., Ramirez, B., Silva, C., Rotellar, F., Cienfuegos, J., Salvador, J. and Frubeck, G. (2011) 'Association of increased Visfatin/PBEF/NAMPT circulating concentrations and gene expression levels in peripheral blood cells with lipid metabolism and fatty liver in human morbid obesity', *Nutrition, Metabolism and Cardiovascular Diseases,* 21, (4), pp. 245-253.

Catalano, P. M., Farrell, K., Thomas, A., Huston-Presley, L., Mencin, P., de Mouzon, S. H. and Amini, S. B. (2009) 'Perinatal risk factors for childhood obesity and metabolic dysregulation', *The American journal of clinical nutrition,* 90, (5), pp. 1303.

Challis, B., Coll, A., Yeo, G., Pinnock, S., Dickson, S., Thresher, R., Dixon, J., Zahn, D., Rochford, J. and White, A. (2004) 'Mice lacking pro-opiomelanocortin are sensitive to high-fat feeding but respond normally to the acute anorectic effects of peptide-YY3-36', *Proceedings of the National Academy of Sciences of the United States of America,* 101, (13), pp. 4695.

Challis, B. G., Pritchard, L. E., Creemers, J. W. M., Delplanque, J., Keogh, J. M., Luan, J., Wareham, N. J., Yeo, G. S. H., Bhattacharyya, S. and Froguel, P. (2002) 'A missense mutation disrupting a dibasic prohormone processing site in pro-opiomelanocortin (POMC) increases susceptibility to early-onset obesity through a novel molecular mechanism', *Human molecular genetics,* 11, (17), pp. 1997.

Chan, D. C., Watts, G. F., Gan, S. K., Wong, A. T. Y., Ooi, E. M. M. and Barrett, P. H. R. (2010) 'Nonalcoholic Fatty Liver Disease as the Transducer of Hepatic Oversecretion of Very-Low-Density Lipoprotein-Apolipoprotein B-100 in Obesity', *Arteriosclerosis, thrombosis, and vascular biology,* 30, (5), pp. 1043-1050.

Chatrath, H., Vuppalanchi, R., Chalasani. (2012), *Semin Liver Dis*, 32, (1)., pp. 22-9.

Chen, C. L., Ip, S. M., Cheng, D., Wong, L. C. and Ngan, H. (2000) 'Loss of imprinting of the IGF-II and H19 genes in epithelial ovarian cancer', *Clinical cancer research,* 6, (2), pp. 474.

Chen, H., Trumbauer, M., Chen, A., Weingarth, D., Adams, J., Frazier, E., Shen, Z., Marsh, D., Feighner, S. and Guan, X. M. (2004) 'Orexigenic action of peripheral ghrelin is mediated by neuropeptide Y and agouti-related protein', *Endocrinology,* 145, (6), pp. 2607.

Chen, H. H., Lee, W. J., Fann, C. S. J., Bouchard, C. and Pan, W. H. (2009a) 'Severe obesity is associated with novel single nucleotide polymorphisms of the ESR1 and PPAR locus in Han Chinese', *The American journal of clinical nutrition,* 90, (2), pp. 255.

Chen, J., Aronow, B. J. and Jegga, A. G. (2009b) 'Disease candidate gene identification and prioritization using protein interaction networks', *BMC bioinformatics,* 10, (1), pp. 73.

Chen, N. C., Yang, F., Capecci, L. M., Gu, Z., Schafer, A. I., Durante, W., Yang, X. F. and Wang, H. (2010) 'Regulation of homocysteine metabolism and methylation in human and mouse tissues', *The FASEB Journal,* 24, (8), pp. 2804.

Cheung, C. C., Clifton, D. K. and Steiner, R. A. (1997) 'Proopiomelanocortin neurons are direct targets for leptin in the hypothalamus', *Endocrinology,* 138, (10), pp. 4489.

Cheung, P., Tanner, K. G., Cheung, W. L., Sassone-Corsi, P., Denu, J. M. and Allis, C. D. (2000) 'Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation', *Molecular Cell,* 5, (6), pp. 905-915.

Chmurzynska, A. (2010) 'Fetal programming: link between early nutrition, DNA methylation, and complex diseases', *Nutrition reviews,* 68, (2), pp. 87-98.

Chowdhury, S., Cleves, M. A., MacLeod, S. L., James, S. J., Zhao, W. and Hobbs, C. A. (2011) 'Maternal DNA hypomethylation and congenital heart defects', *Birth Defects Research,* 91, (2), pp. 69-76.

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. and Bueno, R. (2009) 'Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context', *PLoS genetics,* 5, (8), pp. e1000602.

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

Chumbley, J. R. and Friston, K. J. (2009) 'False discovery rate revisited: FDR and topological inference using Gaussian random fields', *Neuroimage,* 44, (1), pp. 62-70.

Clark, S. J., Harrison, J., Paul, C. L. and Frommer, M. (1994) 'High sensitivity mapping of methylated cytosines', *Nucleic Acids Research,* 22, (15), pp. 2990.

Cnattingius, S., Villamor, E., Lagerros, Y., Wikstrom, A. and Granath, F. (2011) 'High birth weight and obesity: a vicious circle across generations', *International Journal of Obesity*. Epu ahead of print.

Coen, P. M., Dub√©, J. J., Amati, F., Stefanovic-Racic, M., Ferrell, R. E., Toledo, F. G. S. and Goodpaster, B. H. (2010) 'Insulin resistance is associated with higher intramyocellular triglycerides in type I but not type II myocytes concomitant with higher ceramide content', *Diabetes,* 59, (1), pp. 80-88.

Cole, T. J., Bellizzi, M. C., Flegal, K. M. and Dietz, W. H. (2000) 'Establishing a standard definition for child overweight and obesity worldwide: international survey', *Bmj,* 320, (7244), pp. 1240.

Colella, S., Shen, L., Baggerly, K., Issa, J. and Krahe, R. (2003) 'Sensitive and quantitative universal Pyrosequencing,Ñ¢ methylation analysis of CpG sites', *Biotechniques,* 35, (1), pp. 146-151.

Conen, D., Tedrow, U. B., Cook, N. R., Buring, J. E. and Albert, C. M. (2010) 'Birth weight is a significant risk factor for incident atrial fibrillation', *Circulation,* 122, (8), pp. 764-770.

Considine, R. V. (2011) 'Increased Serum Leptin Indicates Leptin Resistance in Obesity', *Clinical Chemistry,* 57, (10), pp. 1461.

Constancia, M., Hemberger, M., Hughes, J., Dean, W., Ferguson-Smith, A., Fundele, R., Stewart, F., Kelsey, G., Fowden, A. and Sibley, C. (2002) 'Placental-specific IGF-II is a major modulator of placental and fetal growth', *Nature,* 417, (6892), pp. 945-948.

Cook, J. D. (2010) 'Frataxin, the mitochondrial iron chaperone for Fe-S cluster bioassembly'.

Coppari, R., Ichinose, M., Lee, C. E., Pullen, A. E., Kenny, C. D., McGovern, R. A., Tang, V., Liu, S. M., Ludwig, T. and Chua Jr, S. C. (2005) 'The hypothalamic arcuate nucleus: a key site for mediating leptin's effects on glucose homeostasis and locomotor activity', *Cell Metabolism,* 1, (1), pp. 63-72.

Cosgrove, G. P., Yang, I., Leach, S. and Schwartz, D. A. (2011) 'Differential Methylation of CpG 'Island Shores' Regulates Gene Expression in Idiopathic Pulmonary Fibrosis', *Proceedings of the American Thoracic Society,* 8, (2), pp. 211-212.

Cowley, M. A., Smart, J. L., Rubinstein, M., Cerd·n, M. G., Diano, S., Horvath, T. L., Cone, R. D. and Low, M. J. (2001) 'Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus', *Nature,* 411, (6836), pp. 480-484.

Cripps, R. L., Martin-Gronert, M. S., Archer, Z. A., Hales, C. N., Mercer, J. G. and Ozanne, S. E. (2009) 'Programming of hypothalamic energy balance gene expression in rats by maternal diet during pregnancy and lactation', *Clinical Science,* 117, (2), pp. 85-93.

Csankovszki, G., Nagy, A. and Jaenisch, R. (2001) 'Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation', *The Journal of cell biology,* 153, (4), pp. 773.

Cui, H., Cruz-Correa, M., Giardiello, F. M., Hutcheon, D. F., Kafonek, D. R., Brandenburg, S., Wu, Y., He, X., Powe, N. R. and Feinberg, A. P. (2003) 'Loss of IGF2 imprinting: a potential marker of colorectal cancer risk', *Science,* 299, (5613), pp. 1753.

Da Wei Huang, B. T. S. and Lempicki, R. A. (2008) 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources', *Nature protocols,* 4, (1), pp. 44-57.

De Carvalho, D. D., You, J. S. and Jones, P. A. (2010) 'DNA methylation and cellular reprogramming', *Trends in cell biology,* 20, (10), pp. 609-617.

Deaton, A. M. and Bird, A. (2011) 'CpG islands and the regulation of transcription', *Genes & Development,* 25, (10), pp. 1010.

Deckert, J. and Struhl, K. (2001) 'Histone acetylation at promoters is differentially affected by specific activators and repressors', *Molecular and cellular biology,* 21, (8), pp. 2726.

DeFronzo, R. A. and Ferrannini, E. (1991) 'Insulin resistance. A multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease', *Diabetes Care,* 14, (3), pp. 173.

Delahaye, F., Breton, C., Risold, P. Y., Enache, M., Dutriez-Casteloot, I., Laborie, C., Lesage, J. and Vieau, D. (2008) 'Maternal perinatal undernutrition drastically reduces postnatal leptin surge and affects the development of arcuate nucleus proopiomelanocortin neurons in neonatal male rat pups', *Endocrinology,* 149, (2), pp. 470.

DeMeo, D., Qiu, W., Bacherman, H., Washko, G. R., Make, B. J., Crapo, J. and Silverman, E. K. (2011) 'Genome-Wide DNA Methylation As A Peripheral Biomarker Of Emphysema', *American journal of respiratory and critical care medicine,* 183, (1 MeetingAbstracts), pp. A4093.

Deng, J., Shoemaker, R., Xie, B., Gore, A., LeProust, E. M., Antosiewicz-Bourget, J., Egli, D., Maherali, N., Park, I. H. and Yu, J. (2009) 'Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming', *Nature Biotechnology,* 27, (4), pp. 353-360.

Dietrich, D., Kneip, C., Raji, O., Liloglou, T., Seegebarth, A., Schlegel, T., Flemming, N., Rausch, S., Distler, J. and Fleischhacker, M. (2011) 'Performance evaluation of the DNA methylation biomarker SHOX2 for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates', *International journal of oncology*, 40, (3), pp. 25-32.

Dietz, W. and Robinson, T. (1998) 'Use of the body mass index (BMI) as a measure of overweight in children and adolescents', *The Journal of pediatrics,* 132, (2), pp. 191.

Dimas, A. S., Deutsch, S., Stranger, B. E., Montgomery, S. B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Arcelus, M. G. and Sekowska, M. (2009) 'Common regulatory variation impacts gene expression in a cell typeñdependent manner', *Science,* 325, (5945), pp. 1246.

Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J. and Loewer, S. (2009) 'Differential methylation of tissue-and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts', *Nature genetics,* 41, (12), pp. 1350-1353.

Dolinoy, D. C., Weidman, J. R. and Jirtle, R. L. (2007) 'Epigenetic gene regulation: linking early developmental environment to adult disease', *Reproductive Toxicology,* 23, (3), pp. 297-307.

Dolinoy, D. C., Weidman, J. R., Waterland, R. A. and Jirtle, R. L. (2006) 'Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome', *Environmental Health Perspectives,* 114, (4), pp. 567.

Down, T. A., Rakyan, V. K., Turner, D. J., Flicek, P., Li, H., Kulesha, E., Gr√§f, S., Johnson, N., Herrero, J. and Tomazou, E. M. (2008) 'A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis', *Nature biotechnology,* 26, (7), pp. 779-785.

Dudziec, E., Miah, S., Choudhry, H. M. Z., Owen, H. C., Blizard, S., Glover, M., Hamdy, F. C. and Catto, J. W. F. (2011) 'Hypermethylation of CpG islands and shores around specific microRNAs and mirtrons is associated with the phenotype and presence of bladder cancer', *Clinical Cancer Research,* 17, (6), pp. 1287.

Dzamko, N., van Denderen, B. J. W., Hevener, A. L., J√∏rgensen, S. B., Honeyman, J., Galic, S., Chen, Z. P., Watt, M. J., Campbell, D. J. and Steinberg, G. R. (2010) 'AMPK Œ≤1 deletion reduces appetite, preventing obesity and hepatic insulin resistance', *Journal of Biological Chemistry,* 285, (1), pp. 115.

Eberle, F. C., Hanson, J. C., Killian, J. K., Wei, L., Ylaya, K., Hewitt, S. M., Jaffe, E. S., Emmert-Buck, M. R. and Rodriguez-Canales, J. (2010) 'Immunoguided laser assisted microdissection techniques for DNA methylation analysis of archival tissue specimens', *Journal of Molecular Diagnostics,* 12, (4), pp. 394.

Eckardt, K., Taube, A. and Eckel, J. (2011) 'Obesity-associated insulin resistance in skeletal muscle: Role of lipid accumulation and physical inactivity', *Reviews in Endocrine & Metabolic Disorders,* 12, (3), pp. 163-172.

Eggenschwiler, J., Ludwig, T., Fisher, P., Leighton, P. A., Tilghman, S. M. and Efstratiadis, A. (1997) 'Mouse mutant embryos overexpressing IGF-II exhibit phenotypic features of the BeckwithñWiedemann and SimpsonñGolabiñBehmel syndromes', *Genes & development,* 11, (23), pp. 3128.

Ekins, S., Nikolsky, Y., Bugrim, A., Kirillov, E. and Nikolskaya, T. (2006) 'Pathway mapping tools for analysis of high content data', *Methods in Molecular Biology,* 356, pp. 319.

Elliott, H., Shah, U., Frayling, T., Xie, W., Walke, M. and Relton, C. (2012) 'Differential methylation of the Type 2 Diabetes susceptibility locus KCNQ1 is associated with insulin sensitivity and is predicted by CpG site-specific genetic variation', *RISC Consortium,* Submitted.

Epel, E. S., Tomiyama, A. J. and Dallman, M. F. (2011) 'Stress and Reward Neural Networks, Eating, and Obesity', in Handbook of food addiction. Oxford: Oxford University Press.

Eriksson, J., Forsen, T., Tuomilehto, J., Jaddoe, V., Osmond, C. and Barker, D. (2002) 'Effects of size at birth and childhood growth on the insulin resistance syndrome in elderly individuals', *Diabetologia,* 45, (3), pp. 342-348.

Fagerberg, B., Bondjers, L. and Nilsson, P. (2004) 'Low birth weight in combination with catch-up growth predicts the occurrence of the metabolic syndrome in men at late middle age: the Atherosclerosis and Insulin Resistance study', *Journal of internal medicine,* 256, (3), pp. 254-259.

Fang, Y. C., Huang, H. C. and Juan, H. F. (2008) 'MeInfoText: associated gene methylation and cancer information from text mining', *BMC bioinformatics,* 9, (1), pp. 22.

Fang, Y. C., Lai, P. T., Dai, H. J. and Hsu, W. L. (2011) 'MeInfoText 2.0: gene methylation and cancer relation extraction from biomedical literature', *BMC bioinformatics,* 12, (1), pp. 471.

Farooqi, I. and O'Rahilly, S. (2007) 'Genetic factors in human obesity', *Obesity Reviews,* 8, pp. 37-40.

Faust, I. M., Johnson, P. R. and Hirsch, J. (1980) 'Long-term effects of early nutritional experience on the development of obesity in the rat', *The Journal of nutrition,* 110, (10), pp. 2027.

Feinberg, A. P. (1999) 'Imprinting of a genomic domain of 11p15 and loss of imprinting in cancer: an introduction', *Cancer research,* 59, (7 Supplement), pp. 1743s.

Feinberg, A. P. (2007) 'Phenotypic plasticity and the epigenetics of human disease', *Nature,* 447, (7143), pp. 433.

Feinberg, A. P. (2008) 'Epigenetics at the epicenter of modern medicine', *JAMA: The Journal of the American Medical Association,* 299, (11), pp. 1345.

Fenech, M., Aitken, C. and Rinaldi, J. (1998) 'Folate, vitamin B12, homocysteine status and DNA damage in young Australian adults', *carcinogenesis,* 19, (7), pp. 1163.

Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Bruning, J. C. and Ruther, U. (2009) 'Inactivation of the Fto gene protects from obesity', *Nature,* 458, (7240), pp. 894-898.

Flom, J. D., Ferris, J. S., Liao, Y., Tehranifar, P., Richards, C. B., Cho, Y. H., Gonzalez, K., Santella, R. M. and Terry, M. B. (2011) 'Prenatal Smoke Exposure and Genomic DNA Methylation in a Multi-ethnic Urban Birth Cohort', *Cancer Epidemiology Biomarkers & Prevention*, 20, (12), pp. 2518-23

Fraga, M. F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P. A. and Esteller, M. (2003) 'The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties', *Nucleic Acids Research,* 31, (6), pp. 1765-1774.

Fraga, M. F. and Esteller, M. (2007) 'Epigenetics and aging: the targets and the marks', *Trends in Genetics,* 23, (8), pp. 413-418.

Friedewald, W. T., Levy, R. I. and Fredrickson, D. S. (1972) 'Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge', *Clin Chem,* 18, (4337382), pp. 499-502.

Frisch, M., Klocke, B., Haltmeier, M. and Frech, K. (2009) 'LitInspector: literature and signal transduction pathway mining in PubMed abstracts', *Nucleic Acids Research,* 37, (suppl 2), pp. W135-W140.

Fryer, A. A., Emes, R. D., Ismail, K. M. K., Haworth, K. E., Mein, C., Carroll, W. D. and Farrell, W. E. (2011) 'Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans', *Epigenetics,* 6, (1), pp. 86.

Fryer, A. A., Nafee, T. M., Ismail, K., Carroll, W. D., Emes, R. D. and Farrell, W. E. (2009) 'LINE-1 DNA methylation is inversely correlated with cord plasma homocysteine in man: a preliminary study', *Epigenetics: Official Journal of the DNA Methylation Society,* 4, (6), pp. 394.

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T. and Nakao, M. (2003) 'Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression', *Journal of Biological Chemistry,* 278, (26), pp. 24132.

Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P. and Kouzarides, T. (2003) 'The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation', *Journal of Biological Chemistry,* 278, (6), pp. 4035.

Gagnon, J. F., Sanschagrin, F., Jacob, S., Tremblay, A. A., Provencher, L., Robert, J. and Morin, C. (2010) 'Quantitative DNA methylation analysis of laser capture microdissected formalin-fixed and paraffin-embedded tissues', *Experimental and molecular pathology,* 88, (1), pp. 184-189.

Gao, H., Falt, S., Sandelin, A., Gustafsson, J. Ö. and Dahlman-Wright, K. (2008) 'Genome-wide identification of estrogen receptor Œ±-binding sites in mouse liver', *Molecular Endocrinology,* 22, (1), pp. 10-22.

Gao, W.-l., Li, D., Xiao, Z.-x., Liao, Q.-p., Yang, H.-x., Li, Y.-x., Ji, L. and Wang, Y.-l. (2011a) 'Detection of global DNA methylation and paternally imprinted H19 gene methylation in preeclamptic placentas', *Hypertension Research - Clinical & Experimental,* 34, (5), pp. 655-61.

Gao, Y., Baccarelli, A., Shu, X., Ji, B., Yu, K., Tarantini, L., Yang, G., Li, H., Hou, L. and Rothman, N. (2011b) 'Blood leukocyte Alu and LINE-1 methylation and gastric cancer risk in the Shanghai Women's Health Study', *British Journal of Cancer*, 106, (3), pp. 585-91.

Gardner, D. S. L., Hosking, J., Metcalf, B. S., Jeffery, A. N., Voss, L. D. and Wilkin, T. J. (2009) 'Contribution of early weight gain to childhood overweight and metabolic health: a longitudinal study (EarlyBird 36)', *Pediatrics,* 123, (1), pp. e67.

Garland, T., Schutz, H., Chappell, M. A., Keeney, B. K., Meek, T. H., Copes, L. E., Acosta, W., Drenowatz, C., Maciel, R. C. and Van Dijk, G. (2011) 'The biological control of voluntary exercise, spontaneous physical activity and daily energy expenditure in relation to obesity: human and rodent perspectives', *Journal of Experimental Biology,* 214, (2), pp. 206.

Gaunt, T. R., Cooper, J. A., Miller, G. J., Day, I. N. M. and OíDell, S. D. (2001) 'Positive associations between single nucleotide polymorphisms in the IGF2 gene region and body mass index in adult males', *Human molecular genetics,* 10, (14), pp. 1491.

Gee, C. E., Chen, C. L. C., Roberts, J. L., Thompson, R. and Watson, S. J. (1983) 'Identification of proopiomelanocortin neurones in rat hypothalamus by in situ cDNA-mRNA hybridization', *Nature*, 306, (5941), pp. 374-6.

Geneletti, S., Richardson, S. and Best, N. (2009) 'Adjusting for selection bias in retrospective, case-control studies', *Biostatistics,* 10, (1), pp. 17-31.

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. and Troncoso, J. (2010) 'Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain', *PLoS genetics,* 6, (5), pp. e1000952.

Ginsberg, H. N. and MacCallum, P. R. (2009) 'The obesity, metabolic syndrome, and type 2 diabetes mellitus pandemic: Part I. Increased cardiovascular disease risk and the importance of atherogenic dyslipidemia in persons with the metabolic syndrome and type 2 diabetes mellitus', *Journal of the cardiometabolic syndrome,* 4, (2), pp. 113-119.

Gloss, B., Patterson, K., Barton, C., Gonzalez, M., Scurry, J., Hacker, N., Sutherland, R., O'Brien, P. and Clark, S. (2011) 'Integrative genome-wide expression and promoter DNA methylation profiling identifies a potential novel panel of ovarian cancer epigenetic biomarkers', *Cancer letters*, 318, (1), pp. 76-85.

Gluckman, P. D. and Hanson, M. A. (2004a) 'Developmental origins of disease paradigm: a mechanistic and evolutionary perspective', *Pediatric Research,* 56, (3), pp. 311.

Gluckman, P. D. and Hanson, M. A. (2004b) 'The developmental origins of the metabolic syndrome', *Trends in Endocrinology and Metabolism,* 15, (4), pp. 183-187.

Gluckman, P. D. and Hanson, M. A. (2006) 'The Developmental Origins of Health and Disease', *Early Life Origins of Health and Disease*, pp. 1-7.

Gluckman, P. D., Hanson, M. A., Cooper, C. and Thornburg, K. L. (2008) 'Effect of *in utero* and early-life conditions on adult health and disease', *New England Journal of Medicine,* 359, (1), pp. 61-73.

Gluckman, P. D., Hanson, M. A. and Spencer, H. G. (2005) 'Predictive adaptive responses and human evolution', *Trends in Ecology & Evolution,* 20, (10), pp. 527-533.

Godfrey, K. M., Lillycrop, K. A., Hanson, M. A. and Burdge, G. C. (2011a) 'Epigenetic mechanisms in the developmental origins of adult disease', *Epigenetic Aspects of Chronic Diseases*, pp. 187-204.

Godfrey, K. M., Sheppard, A., Gluckman, P. D., Lillycrop, K. A., Burdge, G. C., McLean, C., Rodford, J., Slater-Jefferies, J. L., Garratt, E. and Crozier, S. R. (2011b) 'Epigenetic gene

promoter methylation at birth is associated with child's later adiposity', *Diabetes,* 60, (5), pp. 1528-1534.

Gomes, M., Soares, M., Pasqualim-Neto, A., Marcondes, C., LÙbo, R. and Ramos, E. (2005) 'Association between birth weight, body mass index and IGF2/ApaI polymorphism', *Growth hormone & IGF research,* 15, (5), pp. 360-362.

Gomez-Acevedo, H., Shankar, K., Ronis, M. and Badger, T. (2011) 'A methyl-seq analyses of rat offspring liver reveals maternal obesity-induced alterations in dna methylation status at weaning', *ARS,* 3, (2), pp. 23.

Gong, L., Pan, Y. X. and Chen, H. (2010) 'Gestational low protein diet in the rat mediates Igf2 gene expression in male offspring via altered hepatic DNA methylation', *Epigenetics: official journal of the DNA Methylation Society,* 5, (7), pp. 619-626.

Grant, P. A. (2001) 'A tale of histone modifications', *Genome Biology,* 2, (4), pp. REVIEWS0003.

Grill, H. J., Schwartz, M. W., Kaplan, J. M., Foxhall, J. S., Breininger, J. and Baskin, D. G. (2002) 'Evidence that the caudal brainstem is a target for the inhibitory effect of leptin on food intake', *Endocrinology,* 143, (1), pp. 239.

Groom, A., Embleton, N., Korada, M., Swan, D., Cordell, H., Mathers, J. C. and Relton, C. L. (2010) 'Growth in early life is associated with altered gene expression, DNA methylation and body composition in childhood', *Archives of Disease in Childhood,* 95, pp. A53-A54.

Groom, A., Potter, C., Swan, D. C., Fatemifar, G., Evans, D. M., Ring, S. M., Turcot, V., Pearce, M. S., Embleton, N. D. and Smith, G. D. (2012) 'Postnatal Growth and DNA Methylation Are Associated With Differential Gene Expression of the TACSTD2 Gene and Childhood Fat Mass', *Diabetes,* 61, (2), pp. 391-400.

Grunau, C., Clark, S. and Rosenthal, A. (2001) 'Bisulfite genomic sequencing: systematic investigation of critical experimental parameters', *Nucleic Acids Research,* 29, (13), pp. e65.

Gu, H., Bock, C., Mikkelsen, T. S., J√§ger, N., Smith, Z. D., Tomazou, E., Gnirke, A., Lander, E. S. and Meissner, A. (2010) 'Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution', *Nature methods,* 7, (2), pp. 133-136.

Guerrero-Preston, R., Goldman, L. R., Brebi-Mieville, P., Ili-Gangas, C., LeBron, C., Hernandez-Arroyo, M., Witter, F. R., Apelberg, B. J., Roystacher, M. and Jaffe, A. (2010) 'Global DNA hypomethylation is associated with *in utero* exposure to cotinine and perfluorinated alkyl compounds', *Epigenetics,* 5, (6), pp. 539-546.

Guillod-Maximin, E., Roy, A. F., Vacher, C., Aubourg, A., Bailleux, V., Lorsignol, A., Penicaud, L., Parquet, M. and Taouis, M. (2009) 'Adiponectin receptors are expressed in hypothalamus and colocalised with proopiomelanocortin and neuropeptide Y in rodent arcuate neurons', *Journal of Endocrinology,* 200, (1), pp. 93.

Gupta, N., Goel, K., Shah, P. and Misra, A. (2012) 'Childhood Obesity in Developing Countries: Epidemiology, Determinants, and Prevention', *Endocrine Reviews.*

Gurley, L. R., D'Anna, J. A., Barham, S. S., Deaven, L. L. and Tobey, R. A. (1978) 'Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells', *European Journal of Biochemistry,* 84, (1), pp. 1-15.

Gusev, Y., Schmittgen, T., Lerner, M., Postier, R. and Brackett, D. (2007) 'Computational analysis of biological functions and pathways collectively targeted by co-expressed microRNAs in cancer', *BMC bioinformatics,* 8, (Suppl 7), pp. S16.

Haffner, S., Valdez, R., Hazuda, H., Mitchell, B., Morales, P. and Stern, M. (1992) 'Prospective analysis of the insulin-resistance syndrome (syndrome X)', *Diabetes,* 41, (6), pp. 715.

Hales, C. N. and Barker, D. J. (2001) 'The thrifty phenotype hypothesis', *British Medical Bulletin,* 60, pp. 5-20.

Hales, C. N. and Barker, D. J. P. (1992) 'Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis', *Diabetologia,* 35, (7), pp. 595-601.

Hall, J. M., Couse, J. F. and Korach, K. S. (2001) 'The multifaceted mechanisms of estradiol and estrogen receptor signaling', *Journal of Biological Chemistry,* 276, (40), pp. 36869.

Hanson, M. A. and Gluckman, P. D. (2008) 'Developmental origins of health and disease: new insights', *Basic & Clinical Pharmacology & Toxicology,* 102, (2), pp. 90-3.

Haroon, N., Tsui, F. W. L., OíShea, F. D., Chiu, B., Tsui, H. W., Zhang, H., Marshall, K. and Inman, R. D. (2010) 'From gene expression to serum proteins: biomarker discovery in ankylosing spondylitis', *Annals of the rheumatic diseases,* 69, (01), pp. 297.

Hauge, C. and Frodin, M. (2006) 'RSK and MSK in MAP kinase signalling', *Journal of cell science,* 119, (15), pp. 3021.

He, C. (2010) 'Grand Challenge Commentary: RNA epigenetics?', *Nature Chemical Biology,* 6, (12), pp. 863-865.

He, J., Watkins, S. and Kelley, D. E. (2001) 'Skeletal muscle lipid content and oxidative enzyme activity in relation to muscle fiber type in type 2 diabetes and obesity', *Diabetes,* 50, (4), pp. 817.

Heijmans, B. T., Kremer, D., Tobi, E. W., Boomsma, D. I. and Slagboom, P. E. (2007) 'Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus', *Human molecular genetics,* 16, (5), pp. 547.

Heijmans, B. T. and Mill, J. (2012) 'Commentary: The seven plagues of epigenetic epidemiology', *International journal of epidemiology*, 41, (1), pp. 74.

Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., Slagboom, P. E. and Lumey, L. (2008) 'Persistent epigenetic differences associated with prenatal exposure to famine in humans', *Proceedings of the National Academy of Sciences,* 105, (44), pp. 17046.

Henriksen, T. and Clausen, T. (2002) 'The fetal origins hypothesis: placental insufficiency and inheritance versus maternal malnutrition in well-nourished populations', *Acta obstetricia et gynecologica Scandinavica,* 81, (2), pp. 112-114.

Herbstman, J., Tang, D., Zhu, D. and Perera, F. (2009) 'Prenatal exposure to polycyclic aromatic hydrocarbons and CpG methylation', *Epidemiology,* 20, (6), pp. S93.

Herbstman, J. B., Tang, D., Zhu, D., Qu, L., Sjödin, A., Li, Z., Camann, D. and Perera, F. P. (2012) 'Prenatal Exposure to Polycyclic Aromatic Hydrocarbons, Benzo[a]pyrene–DNA Adducts, and Genomic DNA Methylation in Cord Blood', *Environ Health Perspect*, 120, (5).

Herbstman, J. B., Tang, D., Zhu, D., Qu, L., Sj√∂din, A., Li, Z., Camann, D. and Perera, F. P. (2012) 'Prenatal Exposure to Polycyclic Aromatic Hydrocarbons, Benzo [a] Pyrene-DNA Adducts and Genomic DNA Methylation in Cord Blood', *Environmental health perspectives.*

Hernandez, D. G., Nalls, M. A., Gibbs, J. R., Arepalli, S., van der Brug, M., Chong, S., Moore, M., Longo, D. L., Cookson, M. R. and Traynor, B. J. (2011) 'Distinct DNA

methylation changes highly correlated with chronological age in the human brain', *Human molecular genetics,* 20, (6), pp. 1164.

Hewson, A. and Dickson, S. (2000) 'Systemic Administration of Ghrelin Induces Fos and Egr 1 Proteins in the Hypothalamic Arcuate Nucleus of Fasted and Fed Rats', *Journal of Neuroendocrinology,* 12, (11), pp. 1047-1049.

Hinoue, T., Weisenberger, D. J., Lange, C. P. E., Shen, H., Byun, H. M., Van Den Berg, D., Malik, S., Pan, F., Noushmehr, H. and van Dijk, C. M. (2012) 'Genome-scale analysis of aberrant DNA methylation in colorectal cancer', *Genome Research,* 22, (2), pp. 271-282.

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R. and Sasaki, H. (2008) 'Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development', *Genes & development,* 22, (12), pp. 1607.

Hitchins, M. P., Rapkins, R. W., Kwok, C. T., Srivastava, S., Wong, J. J. L., Khachigian, L. M., Polly, P., Goldblatt, J. and Ward, R. L. (2011) 'Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR', *Cancer cell,* 20, (2), pp. 200-213.

Hitchins, M. P., Wong, J. J. L., Suthers, G., Suter, C. M., Martin, D. I. K., Hawkins, N. J. and Ward, R. L. (2007) 'Inheritance of a cancer-associated MLH1 germ-line epimutation', *New England Journal of Medicine,* 356, (7), pp. 697-705.

Hochberg, Z., Feil, R., Constancia, M., Fraga, M., Junien, C., Carel, J. C., Boileau, P., Le Bouc, Y., Deal, C. and Lillycrop, K. (2011) 'Child health, developmental plasticity, and epigenetic programming', *Endocrine Reviews,* 32, (2), pp. 159.

Hofman, P. L., Regan, F., Jackson, W. E., Jefferies, C., Knight, D. B., Robinson, E. M. and Cutfield, W. S. (2004) 'Premature birth and later insulin resistance', *New England Journal of Medicine,* 351, (21), pp. 2179-2186.

Hsiung, D. T., Marsit, C. J., Houseman, E. A., Eddy, K., Furniss, C. S., McClean, M. D. and Kelsey, K. T. (2007) 'Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma', *Cancer Epidemiology Biomarkers & Prevention,* 16, (1), pp. 108-114.

Ibanez, L., Ong, K., Dunger, D. B. and de Zegher, F. (2006) 'Early development of adiposity and insulin resistance after catch-up weight gain in small-for-gestational-age children', *Journal of Clinical Endocrinology & Metabolism,* 91, (6), pp. 2153.

Ibragimova, I., Ibanez de Caceres, I., Hoffman, A. M., Potapova, A., Dulaimi, E., Al-Saleem, T., Hudes, G. R., Ochs, M. F. and Cairns, P. (2010) 'Global reactivation of epigenetically silenced genes in prostate cancer', *Cancer Prevention Research,* 3, (9), pp. 1084.

Ikenasio-Thorpe, B. A., Breier, B. H., Vickers, M. H. and Fraser, M. (2007) 'Prenatal influences on susceptibility to diet-induced obesity are mediated by altered neuroendocrine gene expression', *Journal of endocrinology,* 193, (1), pp. 31.

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R. and Bird, A. P. (2010) 'Orphan CpG islands identify numerous conserved promoters in the mammalian genome', *PLoS genetics,* 6, (9), pp. e1001134.

Inouye, M., Kettunen, J., Soininen, P., Silander, K., Ripatti, S., Kumpula, L. S., Hamalainen, E., Jousilahti, P., Kangas, A. J., Mannisto, S., Savolainen, M. J., Jula, A., Leiviska, J., Palotie, A., Salomaa, V., Perola, M., Ala-Korpela, M. and Peltonen, L. (2010) 'Metabonomic, transcriptomic, and genomic variation of a population cohort', *Mol Syst Biol,* 6, (21179014), pp. 441-441.

Ioshikhes, I. P. and Zhang, M. Q. (2000) 'Large-scale human promoter mapping using CpG islands', *Nature Genetics,* 26, (1), pp. 61-63.

Iqbal, K., Jin, S. G., Pfeifer, G. P. and SzabÛ, P. E. (2011) 'Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine', *Proceedings of the National Academy of Sciences,* 108, (9), pp. 3642.

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M. and Webster, M. (2009) 'The human colon cancer methylome shows similar hypo-and hypermethylation at conserved tissue-specific CpG island shores', *Nature genetics,* 41, (2), pp. 178-186.

Issa, J., Vertino, P. M., Boehm, C. D., Newsham, I. F. and Baylin, S. B. (1996) 'Switch from monoallelic to biallelic human IGF2 promoter methylation during aging and carcinogenesis', *Proceedings of the National Academy of Sciences,* 93, (21), pp. 11757.

Issa, J. P. (2012) 'Age-Related Variation in DNA Methylation', *Epigenetic Epidemiology*, pp. 185-196.

Issa, J. P. J., Ottaviano, Y. L., Celano, P., Hamilton, S. R., Davidson, N. E. and Baylin, S. B. (1994) 'Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon', *Nature genetics,* 7, (4), pp. 536-540.

Jaenisch, R. and Bird, A. (2003) 'Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals', *Nature Genetics,* 33, pp. 245-254.

Jamieson, T., Brizel, D., Killian, J. K., Oka, Y., Jang, H. S., Fu, X., Clough, R., Vollmer, R., Anscher, M. and Jirtle, R. (2003) 'M6P/IGF2R loss of heterozygosity in head and neck cancer associated with poor patient prognosis', *BMC cancer,* 3, (1), pp. 4.

Janovska, A., Hatzinikolas, G., Mano, M. and Wittert, G. A. (2010) 'The effect of dietary fat content on phospholipid fatty acid profile is muscle fiber type dependent', *American Journal of Physiology-Endocrinology And Metabolism,* 298, (4), pp. E779-E786.

Jasik, C. B. and Lustig, R. H. (2008) 'Adolescent obesity and puberty: the 'perfect storm', *Annals of the New York Academy of Sciences,* 1135, (1), pp. 265-279.

Jimenez-Chillaron, J., Hernandez-Valencia, M., Lightner, A., Faucette, R., Reamer, C., Przybyla, R., Ruest, S., Barry, K., Otis, J. and Patti, M. (2006) 'Reductions in caloric intake and early postnatal growth prevent glucose intolerance and obesity associated with low birthweight', *Diabetologia,* 49, (8), pp. 1974-1984.

Jimenez-Chillaron, J. C., Isganaitis, E., Charalambous, M., Gesta, S., Pentinat-Pelegrin, T., Faucette, R. R., Otis, J. P., Chow, A., Diaz, R. and Ferguson-Smith, A. (2009) 'Intergenerational transmission of glucose intolerance and obesity by in utero undernutrition in mice', *Diabetes,* 58, (2), pp. 460-468.

Jirtle, R. L. and Skinner, M. K. (2007) 'Environmental epigenomics and disease susceptibility', *Nature Reviews Genetics,* 8, (4), pp. 253-262.

Jones, A. P., Simson, E. L. and Friedman, M. I. (1984) 'Gestational undernutrition and the development of obesity in rats', *The Journal of nutrition,* 114, (8), pp. 1484.

Jones, J. I. and Clemmons, D. R. (1995) 'Insulin-like growth factors and their binding proteins: biological actions', *Endocrine reviews,* 16, (1), pp. 3-34.

Jones, J. R., Barrick, C., Kim, K. A., Lindner, J., Blondeau, B., Fujimoto, Y., Shiota, M., Kesterson, R. A., Kahn, B. B. and Magnuson, M. A. (2005) 'Deletion of PPAR-γ in adipose tissues of mice protects against high fat diet-induced obesity and insulin resistance', *Proceedings of the National Academy of Sciences of the United States of America,* 102, (17), pp. 6207.

Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J. and Wolffe, A. P. (1998) 'Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription', *Nature Genetics,* 19, (2), pp. 187-191.

Kaijser, M., Edstedt Bonamy, A. K., Akre, O., Cnattingius, S., Granath, F., Norman, M. and Ekbom, A. (2009) 'Perinatal risk factors for diabetes in later life', *Diabetes,* 58, (3), pp. 523.

Kajantie, E., Strang-Karlsson, S., Hovi, P., Raikkonen, K., Pesonen, A. K., Heinonen, K., Jarvenpaa, A. L., Eriksson, J. G. and Andersson, S. (2010) 'Adults born at very low birth weight exercise less than their peers born at term', *The Journal of pediatrics,* 157, (4), pp. 610-616. e1.

Kalra, S. P., Dube, M. G., Sahu, A., Phelps, C. P. and Kalra, P. S. (1991) 'Neuropeptide Y secretion increases in the paraventricular nucleus in association with increased appetite for food', *Proceedings of the National Academy of Sciences,* 88, (23), pp. 10931.

Kalscheuer, V. M., Mariman, E. C., Schepens, M. T., Rehder, H. and Ropers, H. H. (1993) 'The insulinñlike growth factor typeñ2 receptor gene is imprinted in the mouse but not in humans', *Nature genetics,* 5, (1), pp. 74-78.

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R. P., Benes, V., Gannon, F. and Reid, G. (2008) 'Transient cyclical methylation of promoter DNA', *Nature,* 452, (7183), pp. 112-115.

Kannisto, V., Christensen, K. and Vaupel, J. W. (1997) 'No increased mortality in later life for cohorts bom during famine', *American Journal of Epidemiology,* 145, (11), pp. 987.

Kass, S. U., Pruss, D. and Wolffe, A. P. (1997) 'How does DNA methylation repress transcription?', *Trends in Genetics,* 13, (11), pp. 444-449.

Kelley, D. E., McKolanis, T. M., Hegazi, R. A. F., Kuller, L. H. and Kalhan, S. C. (2003) 'Fatty liver in type 2 diabetes mellitus: relation to regional adiposity, fatty acids, and insulin resistance', *American Journal of Physiology-Endocrinology And Metabolism,* 285, (4), pp. E906.

Kestenbaum, B. (2009) 'Cross-Sectional Studies', *Epidemiology and Biostatistics*, pp. 29-31.

Khan, I. Y., Dekou, V., Douglas, G., Jensen, R., Hanson, M. A., Poston, L. and Taylor, P. D. (2005) 'A high-fat diet during rat pregnancy or suckling induces cardiovascular dysfunction in adult offspring', *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* 288, (1), pp. R127.

Khan, I. Y., Taylor, P. D., Dekou, V., Seed, P. T., Lakasing, L., Graham, D., Dominiczak, A. F., Hanson, M. A. and Poston, L. (2003) 'Gender-linked hypertension in offspring of lard-fed pregnant rats', *Hypertension,* 41, (1), pp. 168-175.

Khatri, P. and Ghici, S. (2005) 'Ontological analysis of gene expression data: current tools, limitations, and open problems', *Bioinformatics,* 21, (18), pp. 3587.

Kim, M., Long, T. I., Arakawa, K., Wang, R., Mimi, C. Y. and Laird, P. W. (2010) 'DNA methylation as a biomarker for cardiovascular disease risk', *PloS one,* 5, (3), pp. e9692.

Kim, T. Y., Zhong, S., Fields, C. R., Kim, J. H. and Robertson, K. D. (2006) 'Epigenomic profiling reveals novel and frequent targets of aberrant DNA methylation-mediated silencing in malignant glioma', *Cancer research,* 66, (15), pp. 7490.

Kirk, S. L., Samuelsson, A. M., Argenton, M., Dhonye, H., Kalamatianos, T., Poston, L., Taylor, P. D. and Coen, C. W. (2009) 'Maternal obesity induced by diet in rats permanently influences central processes regulating food intake in offspring', *PLoS One,* 4, (6), pp. e5870.

Kloting, N., Schleinitz, D., Ruschke, K., Berndt, J., Fasshauer, M., Tonjes, A., Schon, M., Kovacs, P., Stumvoll, M. and Bluher, M. (2008) 'Inverse relationship between obesity and FTO gene expression in visceral adipose tissue in humans', *Diabetologia,* 51, (4), pp. 641-647.

Köhler, S., Bauer, S., Horn, D. and Robinson, P. N. (2008) 'Walking the interactome for prioritization of candidate disease genes', *The American Journal of Human Genetics,* 82, (4), pp. 949-958.

Kong, F. M., Anscher, M. S., Washington, M. K., Killian, J. K. and Jirtle, R. L. (2000) 'M6P/IGF2R is mutated in squamous cell carcinoma of the lung', *Oncogene,* 19, (12), pp. 1572-1578.

Krallinger, M., Leitner, F. and Valencia, A. (2010) 'Analysis of biological processes and diseases using text mining approaches', *Methods in Molecular Biology,* 593, pp. 341-382.

Krallinger, M., Valencia, A. and Hirschman, L. (2008) 'Linking genes to literature: text mining, information extraction, and retrieval applications for biology', *Genome Biol,* 9, (Suppl 2), pp. S8.

Kriaucionis, S. and Heintz, N. (2009) 'The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain', *Science,* 324, (5929), pp. 929.

Kubota, N., Terauchi, Y., Miki, H., Tamemoto, H., Yamauchi, T., Komeda, K., Satoh, S., Nakano, R., Ishii, C. and Sugiyama, T. (1999) 'PPAR [gamma] mediates high-fat diet-induced adipocyte hypertrophy and insulin resistance', *Molecular Cell,* 4, (4), pp. 597-609.

Kuh, D., Ben-Shlomo, Y., Lynch, J., Hallqvist, J. and Power, C. (2003) 'Life course epidemiology', *Journal of Epidemiology and Community Health,* 57, (10), pp. 778.

Kuhlow, D., Zarse, K., Voigt, A., Schulz, T. J., Petzke, K. J., Schomburg, L., Pfeiffer, A. F. H. and Ristow, M. (2010) 'Opposing effects of dietary sugar and saturated fat on cardiovascular risk factors and glucose metabolism in mitochondrially impaired mice', *European journal of nutrition,* 49, (7), pp. 417-427.

Kuo, M. H. and Allis, C. D. (1998) 'Roles of histone acetyltransferases and deacetylases in gene regulation', *Bioessays,* 20, (8), pp. 615-626.

Laird, N. M. and Lange, C. (2006) 'Family-based designs in the age of large-scale gene-association studies', *Nature Reviews Genetics,* 7, (5), pp. 385-394.

Laird, P. W. (2010) 'Principles and challenges of genome-wide DNA methylation analysis', *Nature Reviews Genetics,* 11, (3), pp. 191-203.

Lam, T. K. T., Schwartz, G. J. and Rossetti, L. (2005) 'Hypothalamic sensing of fatty acids', *Nature neuroscience,* 8, (5), pp. 579-584.

Langley-Evans, S. C. (2001) 'Fetal programming of cardiovascular function through exposure to maternal undernutrition', *Proceedings of the Nutrition Society,* 60, (4), pp. 505-13

Larijani, M., Frieder, D., Sonbuchner, T. M., Bransteitter, R., Goodman, M. F., Bouhassira, E. E., Scharff, M. D. and Martin, A. (2005) 'Methylation protects cytidines from AID-mediated deamination', *Molecular immunology,* 42, (5), pp. 599-604.

Latos, P. A., Stricker, S. H., Steenpass, L., Pauler, F. M., Huang, R., Senergin, B. H., Regha, K., Koerner, M. V., Warczok, K. E. and Unger, C. (2009) 'An in vitro ES cell imprinting model shows that imprinted expression of the Igf2r gene arises from an allele-specific expression bias', *Development,* 136, (3), pp. 437.

Lau, M., Stewart, C., Liu, Z., Bhatt, H., Rotwein, P. and Stewart, C. L. (1994) 'Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality', *Genes & Development,* 8, (24), pp. 2953.

Lazzer, S., Bedogni, G., Lafortuna, C. L., Marazzi, N., Busti, C., Galli, R., De Col, A., Agosti, F. and Sartorio, A. (2009) 'Relationship between basal metabolic rate, gender, age, and body composition in 8,780 white obese subjects', *Obesity,* 18, (1), pp. 71-78.

Lecker, S. H., Goldberg, A. L. and Mitch, W. E. (2006) 'Protein degradation by the ubiquitin-proteasome pathway in normal and disease states', *Journal of the American Society of Nephrology,* 17, (7), pp. 1807-1819.

Leinninger, G. M., Jo, Y. H., Leshan, R. L., Louis, G. W., Yang, H., Barrera, J. G., Wilson, H., Opland, D. M., Faouzi, M. A. and Gong, Y. (2009) 'Leptin acts via leptin receptor-expressing lateral hypothalamic neurons to modulate the mesolimbic dopamine system and suppress feeding', *Cell metabolism,* 10, (2), pp. 89-98.

Li, E., Bestor, T. H. and Jaenisch, R. (1992) 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality', *Cell,* 69, (6), pp. 915-926.

Li, L. C., Okino, S. T. and Dahiya, R. (2004) 'DNA methylation in prostate cancer', *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer,* 1704, (2), pp. 87-102.

Li, S., Zhao, J. H., Luan, J., Luben, R. N., Rodwell, S. A., Khaw, K. T., Ong, K. K., Wareham, N. J. and Loos, R. J. F. (2010a) 'Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies', *The American Journal of Clinical Nutrition,* 91, (1), pp. 184-190.

Li, Y., He, Y., Qi, L., Jaddoe, V. W., Feskens, E. J. M., Yang, X., Ma, G. and Hu, F. B. (2010b) 'Exposure to the Chinese famine in early life and the risk of hyperglycemia and type 2 diabetes in adulthood', *Diabetes,* 59, (10), pp. 2400.

Li, Y., Jaddoe, V. W., Qi, L., He, Y., Wang, D., Lai, J., Zhang, J., Fu, P., Yang, X. and Hu, F. B. (2011) 'Exposure to the Chinese famine in early life and the risk of metabolic syndrome in adulthood', *Diabetes Care,* 34, (4), pp. 1014-1018.

Li, Y., Liu, L. and Tollefsbol, T. O. (2010c) 'Glucose restriction can extend normal cell lifespan and impair precancerous cell growth through epigenetic control of hTERT and p16 expression', *FASEB J,* 24, (20019239), pp. 1442-1453.

Lillycrop, K. A. and Burdge, G. C. (2010) 'Epigenetic changes in early life and future risk of obesity', *International Journal of Obesity,* 35, (1), pp. 72-83.

Lillycrop, K. A., Slater-Jefferies, J. L., Hanson, M. A., Godfrey, K. M., Jackson, A. A. and Burdge, G. C. (2007) 'Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications', *British Journal of Nutrition,* 97, (06), pp. 1064-1073.

Lin, R. K., Hsu, H. S., Chang, J. W., Chen, C. Y., Chen, J. T. and Wang, Y. C. (2007) 'Alteration of DNA methyltransferases contributes to 5'CpG methylation and poor prognosis in lung cancer', *Lung Cancer,* 55, (2), pp. 205-213.

Liu, F., Killian, J. K., Yang, M., Walker, R. L., Hong, J. A., Zhang, M., Davis, S., Zhang, Y., Hussain, M., Xi, S., Rao, M., Meltzer, P. A. and Schrump, D. S. (2010a) 'Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate', *Oncogene,* 29, (20440268), pp. 3650-3664.

Liu, L., van Groen, T., Kadish, I., Li, Y., Wang, D., James, S. R., Karpf, A. R. and Tollefsbol, T. O. (2011) 'Insufficient DNA methylation affects healthy aging and promotes age-related health problems', *Clinical Epigenetics*, pp. 1-12.

Liu, S., Jones, R. N. and Glymour, M. M. (2010b) 'Implications of lifecourse epidemiology for research on determinants of adult disease', *Public Health Reviews,* 32, pp. 489-511.

Lo, L., McLennan, S. V., Williams, P. F., Bonner, J., Chowdhury, S., McCaughan, G. W., Gorrell, M. D., Yue, D. K. and Twigg, S. M. (2011) 'Diabetes is a progression factor for hepatic fibrosis in a high fat fed mouse obesity model of non-alcoholic steatohepatitis', *Journal of hepatology,* 55, (2), pp. 435-444.

Loos, R. J. F., Fagard, R., Beunen, G., Derom, C. and Vlietinck, R. (2001) 'Birth Weight and Blood Pressure in Young Adults', *Circulation,* 104, (14), pp. 1633-1638.

Lucifero, D., Mertineit, C., Clarke, H. J., Bestor, T. H. and Trasler, J. M. (2002) 'Methylation dynamics of imprinted genes in mouse germ cells', *Genomics,* 79, (4), pp. 530-538.

Lupien, M., Eeckhoute, J., Meyer, C. A., Wang, Q., Zhang, Y., Li, W., Carroll, J. S., Liu, X. S. and Brown, M. (2008) 'FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription', *Cell,* 132, (6), pp. 958-970.

Ma, D. K., Guo, J. U., Ming, G. and Song, H. (2011) 'Bridging Environment and DNA: Activity-Induced Epigenetic Modification in the Adult Brain', *Two Faces of Evil: Cancer and Neurodegeneration*, pp. 113-123.

Ma, H., Baumann, C. T., Li, H., Strahl, B. D., Rice, R., Jelinek, M. A., Aswad, D. W., Allis, C. D., Hager, G. L. and Stallcup, M. R. (2001) 'Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter', *Current Biology,* 11, (24), pp. 1981-1985.

Machin, G. (2009) 'Non-identical monozygotic twins, intermediate twin types, zygosity testing, and the non-random nature of monozygotic twinning: a review', American Journal of Medical Genetics, Part C, *Seminars in Medical Genetics*. 151C, (2), pp. 110-27.

Maegawa, S., Hinkal, G., Kim, H. S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L. A. and Issa, J. P. J. (2010) 'Widespread and tissue specific age-related DNA methylation changes in mice', *Genome Research,* 20, (3), pp. 332-340.

Marobbio, C. M. T., Pisano, I., Porcelli, V., Lasorsa, F. M. and Palmieri, L. (2011) 'Rapamycin reduces oxidative stress in frataxin deficient yeast cells', *Mitochondrion*, 12, (1), pp. 156-61.

Marsit, C. J., Koestler, D. C., Christensen, B. C., Karagas, M. R., Houseman, E. A. and Kelsey, K. T. (2011) 'DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer', *J Clin Oncol,* 29, (9), pp. 1133-9.

Martin-Gronert, M. S. and Ozanne, S. E. (2010) 'Mechanisms linking suboptimal early nutrition and increased risk of type 2 diabetes and obesity', *The Journal of nutrition,* 140, (3), pp. 662.

Martinez-Gonzalez, M., Alfredo Martinez, J., Hu, F., Gibney, M. and Kearney, J. (1999) 'Physical inactivity, sedentary lifestyle and obesity in the European Union', *International Journal of Obesity,* 23, (11), pp. 1192-1201.

Mathers, J. C. (2008) 'Session 2: Personalised nutrition. Epigenomics: a basis for understanding individual differences?', *Proc Nutr Soc,* 67, (18847515), pp. 390-394.

Mathers, J. C. and McKay, J. A. (2009) 'Epigenetics: potential contribution to fetal programming', *Early Nutrition Programming and Health Outcomes in Later Life*, pp. 119-123.

Mathers, J. C., Strathdee, G. and Relton, C. L. (2010) 'Induction of epigenetic alterations by dietary and other environmental factors', *Adv Genet,* 71, (20933124), pp. 3-39.

Matthews, D., Hosker, J., Rudenski, A., Naylor, B., Treacher, D. and Turner, R. (1985) 'Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man', *Diabetologia,* 28, (7), pp. 412-419.

May, O. (2010) 'DNA Methylation: Fingerprints of the (epi) genome'. *Caymanchem* 23, (2), pp. 212.

Mayer-Davis, E. J., Rifas-Shiman, S. L., Zhou, L., Hu, F. B., Colditz, G. A. and Gillman, M. W. (2006) 'Breast-feeding and risk for childhood obesity', *Diabetes Care,* 29, (10), pp. 2231.

McKay, J., Adriaens, M., Ford, D., Relton, C., Evelo, C. T. A. and Mathers, J. (2008) 'Bioinformatic interrogation of expression array data to identify nutritionally regulated genes potentially modulated by DNA methylation', *Genes & Nutrition,* 3, (3), pp. 167-171.

McKay, J., Williams, E. and Mathers, J. (2004) 'Folate and DNA methylation during in utero development and aging', *Biochemical Society Transactions,* 32, (6), pp. 1006-1007.

McKay, J. A., Williams, E. A. and Mathers, J. C. (2011a) 'Effect of Maternal and Post-Weaning Folate Supply on Gene-Specific DNA Methylation in the Small Intestine of Weaning and Adult Apc+/Min and Wild Type Mice', *Frontiers in Epigenomics,* 2.

McKay, J. A., Wong, Y. K., Relton, C. L., Ford, D. and Mathers, J. C. (2011b) 'Maternal folate supply and sex influence gene-specific DNA methylation in the fetal gut', *Molecular Nutrition & Food Research*, 55, (11) pp. 1717-23.

McKay, J. A., Xie, L., Harris, S., Wong, Y. K., Ford, D. and Mathers, J. C. (2011c) 'Blood as a surrogate marker for tissue-specific DNA methylation and changes due to folate depletion in post-partum female mice', *Molecular Nutrition & Food Research*, 55, (7) pp. 1026-35

McMullen, S. and Mostyn, A. (2009) 'Animal models for the study of the developmental origins of health and disease', *Proc Nutr Soc,* 68, pp. 306-320.

Mertens, A. and Holvoet, P. (2001) 'Oxidized LDL and HDL: antagonists in atherothrombosis', *The FASEB Journal,* 15, (12), pp. 2073.

Michels, K. B., Harris, H. R. and Barault, L. (2011) 'Birthweight, Maternal Weight Trajectories and Global DNA Methylation of LINE-1 Repetitive Elements', *PloS one,* 6, (9), pp. e25254.

Milagro, F. I., Campion, J., Cordero, P., Goyenechea, E., Gomez-Uriz, A. M., Abete, I., Zulet, M. A. and Martinez, J. A. (2011) 'A dual epigenomic approach for the search of obesity biomarkers: DNA methylation in relation to diet-induced weight loss', *The FASEB Journal,* 25, (4), pp. 1378-1389.

Miller, E. R., Erlinger, T. P., Young, D. R., Jehn, M., Charleston, J., Rhodes, D., Wasan, S. K. and Appel, L. J. (2002) 'Results of the diet, exercise, and weight loss intervention trial (DEW-IT)', *Hypertension,* 40, (5), pp. 612-618.

Mitrou, P., Boutati, E., Lambadiari, V., Maratou, E., Papakonstantinou, A., Komesidou, V., Sidossis, L., Tountas, N., Katsilambros, N. and Economopoulos, T. (2009) 'Rates of glucose uptake in adipose tissue and muscle in vivo after a mixed meal in women with morbid obesity', *Journal of Clinical Endocrinology & Metabolism,* 94, (8), pp. 2958.

Montague, C. T., Farooqi, I. S., Whitehead, J. P., Soos, M. A., Rau, H., Wareham, N. J., Sewter, C. P., Digby, J. E., Mohammed, S. N. and Hurst, J. A. (1997) 'Congenital leptin deficiency is associated with severe early-onset obesity in humans', *Nature,* 387, (6636), pp. 903-908.

Moore, L. E., Pfeiffer, R. M., Poscablo, C., Real, F. X., Kogevinas, M., Silverman, D., Garcia-Closas, R., Chanock, S., Tardon, A. and Serra, C. (2008) 'Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study', *The lancet oncology,* 9, (4), pp. 359-366.

Moreno-Cermeno, A., Obis, BellÌ, G., Cabiscol, E., Ros, J. and Tamarit, J. (2010) 'Frataxin Depletion in Yeast Triggers Up-regulation of Iron Transport Systems before Affecting Iron-Sulfur Enzyme Activities', *Journal of Biological Chemistry,* 285, (53), pp. 41653.

Movassagh, M., Choy, M. K., Goddard, M., Bennett, M. R., Down, T. A. and Roger, S. Y. F. (2010) 'Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure', *PLoS One,* 5, (1), pp. e8564.

Murgatroyd, C., Patchev, A. V., Wu, Y., Micale, V., Bockm√ºhl, Y., Fischer, D., Holsboer, F., Wotjak, C. T., Almeida, O. F. X. and Spengler, D. (2009) 'Dynamic DNA methylation programs persistent adverse effects of early-life stress', *Nature neuroscience,* 12, (12), pp. 1559-1566.

Musatov, S., Chen, W., Pfaff, D. W., Mobbs, C. V., Yang, X. J., Clegg, D. J., Kaplitt, M. G. and Ogawa, S. (2007) 'Silencing of estrogen receptor in the ventromedial nucleus of hypothalamus leads to metabolic syndrome', *Proceedings of the National Academy of Sciences,* 104, (7), pp. 2501.

Myers, M. G., Leibel, R. L., Seeley, R. J. and Schwartz, M. W. (2010) 'Obesity and leptin resistance: distinguishing cause from effect', *Trends in Endocrinology & Metabolism,* 21, (11), pp. 643-651.

Myslobodsky, M. (2008) 'Ingenuity pathway analysis of clozapine-induced obesity', *Onkologie,* 31, pp. 93-102.

Nakayama, K. I. and Nakayama, K. (2006) 'Ubiquitin ligases: cell-cycle control and cancer', *Nature Reviews Cancer,* 6, (5), pp. 369-381.

Nakkuntod, J., Avihingsanon, Y., Mutirangura, A. and Hirankarn, N. (2011) 'Hypomethylation of LINE-1 but not Alu in lymphocyte subsets of systemic lupus erythematosus patients', *Clinica Chimica Acta,* 412, (15-16), pp. 1457-61.

Naukkarinen, J., Rissanen, A., Kaprio, J. and Pietilainen, K. (2011) 'Causes and consequences of obesity: the contribution of recent twin studies', *International Journal of Obesity*.

Ndumele, C. E., Nasir, K., Conceiˆ√ßao, R. D., Carvalho, J. A. M., Blumenthal, R. S. and Santos, R. D. (2011) 'Hepatic steatosis, obesity, and the metabolic syndrome are independently and additively associated with increased systemic inflammation', *Arteriosclerosis, thrombosis, and vascular biology,* 31, (8), pp. 1927-1932.

Nieman, K. M. and Schalinske, K. L. (2011) 'Insulin administration abrogates perturbation of methyl group and homocysteine metabolism in streptozotocin-treated type 1 diabetic rats', *American Journal of Physiology-Endocrinology And Metabolism*, 301, (3), pp. E560-5.

Niswender, K. D., Morrison, C. D., Clegg, D. J., Olson, R., Baskin, D. G., Myers, M. G., Seeley, R. J. and Schwartz, M. W. (2003) 'Insulin activation of phosphatidylinositol 3-kinase in the hypothalamic arcuate nucleus', *Diabetes,* 52, (2), pp. 227.

Nivoit, P., Morens, C., Van Assche, F., Jansen, E., Poston, L., Remacle, C. and Reusens, B. (2009) 'Established diet-induced obesity in female rats leads to offspring hyperphagia, adiposity and insulin resistance', *Diabetologia,* 52, (6), pp. 1133-1142.

Novakovic, B., Yuen, R., Gordon, L., Penaherrera, M., Sharkey, A., Moffett, A., Craig, J., Robinson, W. and Saffery, R. (2011) 'Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors', *BMC genomics,* 12, (1), pp. 529.

Novikova, S. I., He, F., Bai, J., Cutrufello, N. J., Lidow, M. S. and Undieh, A. S. (2008) 'Maternal cocaine administration in mice alters DNA methylation and gene expression in hippocampal neurons of neonatal and prepubertal offspring', *PloS one,* 3, (4), pp. e1919.

Nyholm, B., Qu, Z., Kaal, A., Pedersen, S. B., Gravholt, C. H., Andersen, J. L., Saltin, B. and Schmitz, O. (1997) 'Evidence of an increased number of type IIb muscle fibers in insulin-resistant first-degree relatives of patients with NIDDM', *Diabetes,* 46, (11), pp. 1822-1828.

Oates, A. J., Schumaker, L. M., Jenkins, S. B., Pearce, A. A., DaCosta, S. A., Arun, B. and Ellis, M. J. C. (1998) 'The mannose 6-phosphate/insulin-like growth factor 2 receptor (M6P/IGF2R), a putative breast tumor suppressor gene', *Breast cancer research and treatment,* 47, (3), pp. 269-281.

Ogino, S., Nosho, K., Kirkner, G. J., Kawasaki, T., Chan, A. T., Schernhammer, E. S., Giovannucci, E. L. and Fuchs, C. S. (2008) 'A cohort study of tumoral LINE-1 hypomethylation and prognosis in colon cancer', *J Natl Cancer Inst,* 100, (19033568), pp. 1734-1738.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999) 'DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development', *Cell,* 99, (3), pp. 247-257.

Ong, K. K. L., Ahmed, M. L., Emmett, P. M., Preece, M. A. and Dunger, D. B. (2000) 'Association between postnatal catch-up growth and obesity in childhood: prospective cohort study', *Bmj,* 320, (7240), pp. 967-971.

Ong, Z. and Muhlhausler, B. (2011) 'Maternal ìjunk-foodî feeding of rat dams alters food choices and development of the mesolimbic reward pathway in the offspring', *The FASEB Journal,* 25, (7), pp. 2167.

Ongenaert, M. and Van Criekinge, W. (2005) 'CpG island characterisation', *Computational Biology Madrid 05-Programme and abstract book*, pp. 92.

Oosting, A., Kegler, D., Boehm, G., Jansen, H. T., van de Heijning, B. J. M. and van der Beek, E. M. (2010) 'N-3 long-chain polyunsaturated fatty acids prevent excessive fat deposition in adulthood in a mouse model of postnatal nutritional programming', *Pediatric Research,* 68, (6), pp. 494-499.

Ozanne, S. E. and Hales, C. N. (2004) 'Lifespan: catch-up growth and obesity in male mice', *Nature,* 427, (6973), pp. 411-412.

Ozanne, S. E., Lewis, R., Jennings, B. J. and Hales, C. N. (2004) 'Early programming of weight gain in mice prevents the induction of obesity by a highly palatable diet', *Clinical Science,* 106, (2), pp. 141-146.

Palacios, D., Summerbell, D., Rigby, P. W. J. and Boyes, J. (2010) 'Interplay between DNA methylation and transcription factor availability: implications for developmental activation of the mouse myogenin gene', *Molecular and cellular biology,* 30, (15), pp. 3805.

Palmke, N., Santacruz, D. and Walter, J. (2011) 'Comprehensive analysis of DNA-methylation in mammalian tissues using MeDIP-chip', *Methods,* 53, (2), pp. 175-184.

Palou, M., Picó, C., McKay, J., Sanchez, J., Priego, T., Mathers, J. and Palou, A. (2011) 'Protective effects of leptin during the suckling period against later obesity may be associated with changes in promoter methylation of the hypothalamic pro-opiomelanocortin gene', *British Journal of Nutrition,* 106, (5), pp. 769-778.

Park, K., Ciaraldi, T., Abrams-Carter, L., Mudaliar, S., Nikoulina, S. and Henry, R. (1997) 'PPAR-gamma gene expression is elevated in skeletal muscle of obese and type II diabetic subjects', *Diabetes,* 46, (7), pp. 1230.

Parkinson, K. N., Pearce, M. S., Dale, A., Reilly, J. J., Drewett, R. F., Wright, C. M., Relton, C. L., McArdle, P., Le Couteur, A. S. and Adamson, A. J. (2011) 'Cohort Profile: The Gateshead Millennium Study', *International journal of epidemiology,* 40, (2), pp. 308.

Parsons, T., Power, C., Logan, S. and Summerbell, C. (1999) 'Childhood predictors of adult obesity: a systematic review', *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity,* 23, pp. S1.

Patel, P. I. and Isaya, G. (2001) 'Friedreich Ataxia: From GAA TripletñRepeat Expansion to Frataxin Deficiency', *American journal of human genetics,* 69, (1), pp. 15.

Payne, S. R., Serth, J., Schostak, M., Kamradt, J., Strauss, A., Thelen, P., Model, F., Day, J. K., Liebenberg, V. and Morotti, A. (2009) 'DNA methylation biomarkers of prostate cancer: Confirmation of candidates and evidence urine is the most sensitive body fluid for non-invasive detection', *The Prostate,* 69, (12), pp. 1257-1269.

Pearce, M. S., Mann, K. D., Relton, C. L., Francis, R. M., Steele, J. G., Craft, A. W. and 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, (1), pp. 23-8.

Pearce, M. S., Unwin, N. C., Parker, L. and Alberti, K. G. M. M. (2006) 'Life course determinants of insulin secretion and sensitivity at age 50 years: the Newcastle thousand families study', *Diabetes Metab Res Rev,* 22, (15977305), pp. 118-125.

Pearce, M. S., Unwin, N. C., Parker, L. and Craft, A. W. (2009) 'Cohort profile: the Newcastle Thousand Families 1947 birth cohort', *Int J Epidemiol,* 38, (18782900), pp. 932-937.

Pearce, M. S., Unwin, N. C., Relton, C. L., Alberti, K. G. M. M. and Parker, L. (2005) 'Lifecourse determinants of fasting and post-challenge glucose at age 50 years: the Newcastle Thousand Families Study', *Eur J Epidemiol,* 20, (16284869), pp. 915-923.

Perera, F., Tang, W., Herbstman, J., Tang, D., Levin, L., Miller, R. and Ho, S. (2009) 'Relation of DNA methylation of 5'-CpG island of ACSL3 to transplacental exposure to airborne polycyclic aromatic hydrocarbons and childhood asthma', *PloS one,* 4, (2), pp. e4488.

Petry, C. J., Dorling, M., Pawlak, D., Ozanne, S. and Hales, C. (2001) 'Diabetes in old male offspring of rat dams fed a reduced protein diet', *International journal of diabetes research,* 2, pp. 139-143.

Pietiläinen, K., Kannisto, K., Korsheninnikova, E., Rissanen, A., Kaprio, J., Ehrenborg, E., Hamsten, A. and Yki-Järvinen, H. (2006) 'Acquired obesity increases CD68 and tumor necrosis factor-alpha and decreases adiponectin gene expression in adipose tissue: a study in monozygotic twins', *J Clin Endocrinol Metab. ,* 91, (7), pp. 2776-81.

Pilsner, J. R., Hall, M., Liu, X., Ilievski, V., Slavkovich, V., Ahsan, H., Factor-Litvak, P., Graziano, J. and Gamble, M. (2009a) 'Influence of Folic Acid Supplementation on Genomic DNA Methylation: Findings from a Randomised, Placebo-Controlled Trial in Bangladeshi Adults', *Epidemiology,* 20, (6), pp. S214.

Pilsner, J. R., Hu, H., Ettinger, A., Sanchez, B. N., Wright, R. O., Cantonwine, D., Lazarus, A., Lamadrid-Figueroa, H., Mercado-Garcia, A. and Tellez-Rojo, M. M. (2009b) 'Influence of prenatal lead exposure on genomic methylation of cord blood DNA', *Environmental Health Perspectives,* 117, (9), pp. 1466.

Pinnick, K. E. and Karpe, F. (2010) 'DNA methylation of genes in adipose tissue', *The Proceedings of the Nutrition Society*, pp. 1.

Plagemann, A., Harder, T., Brunn, M., Harder, A., Roepke, K., Wittrock-Staar, M., Ziska, T., Schellong, K., Rodekamp, E. and Melchior, K. (2009) 'Hypothalamic proopiomelanocortin promoter methylation becomes altered by early overfeeding: an

epigenetic model of obesity and the metabolic syndrome', *The Journal of Physiology,* 587, (20), pp. 4963.

Poirier, P., Giles, T. D., Bray, G. A., Hong, Y., Stern, J. S., Pi-Sunyer, F. X. and Eckel, R. H. (2006) 'Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss', *Circulation,* 113, (6), pp. 898-918.

Pomplun, D., Voigt, A., Schulz, T. J., Thierbach, R., Pfeiffer, A. F. and Ristow, M. (2007) 'Reduced expression of mitochondrial frataxin in mice exacerbates diet-induced obesity', *Proceedings of the National Academy of Sciences,* 104, (15), pp. 6377.

Ponsonby, A. L., Pezic, A., Cochrane, J., Cameron, F. J., Pascoe, M., Kemp, A. and Dwyer, T. (2011) 'Infant anthropometry, early life infection, and subsequent risk of type 1 diabetes mellitus: a prospective birth cohort study', *Pediatric Diabetes*, 12, (4) pp. 313-21.

Poulsen, P., Vaag, A., Kyvik, K., Jensen, D. M. and Beck-Nielsen, H. (1997) 'Low birth weight is associated with NIDDM in discordant monozygotic and dizygotic twin pairs', *Diabetologia,* 40, (4), pp. 439-446.

Prior, L. J., Head, G. A. and Armitage, J. A. (2011) 'Experimental Models of Maternal Obesity and High-Fat Diet During Pregnancy and Programmed Obesity in the Offspring', *Obesity Before Birth*, pp. 241-259.

Pritchard, L., Turnbull, A. and White, A. (2002) 'Pro-opiomelanocortin processing in the hypothalamus: impact on melanocortin signalling and obesity', *Journal of endocrinology,* 172, (3), pp. 411.

Qi, Y., Henry, B. A., Oldfield, B. J. and Clarke, I. J. (2010) 'The action of leptin on appetite-regulating cells in the ovine hypothalamus: Demonstration of direct action in the absence of the arcuate nucleus', *Endocrinology,* 151, (5), pp. 2106.

Qi, Y., Takahashi, N., Hileman, S. M., Patel, H. R., Berg, A. H., Pajvani, U. B., Scherer, P. E. and Ahima, R. S. (2004) 'Adiponectin acts in the brain to decrease body weight', *Nature medicine,* 10, (5), pp. 524-529.

Qiu, W., Baccarelli, A., Carey, V. J., Boutaoui, N., Bacherman, H., Klanderman, B., Rennard, S., Agusti, A., Anderson, W. and Lomas, D. A. (2011) 'Variable DNA Methylation is Associated with COPD and Lung Function', *American journal of respiratory and critical care medicine*. 185 (4) pp. 373-381.

Rabinovich, E. I., Yakhini, Z., Steinfeld, I., Pandit, K., Yu, G., We, W., Chensny, L. J. and Kaminski, N. (2010) 'The human CpG islands arrays reveal changes in global methylation patterns in idiopathic pulmonary fibrosis', *American Journal of Respiratory and Critical Care Medicine,* 181, (1 MeetingAbstracts), pp. A2017.

Rakyan, V. K., Preis, J., Morgan, H. D. and Whitelaw, E. (2001) 'The marks, mechanisms and memory of epigenetic states in mammals', *Biochem J,* 356, (11336630), pp. 1-10.

Ramirez, L., Zabaleta, E. J. and Lamattina, L. (2010) 'Nitric oxide and frataxin: two players contributing to maintain cellular iron homeostasis', *Annals of botany,* 105, (5), pp. 801.

Ramos, E., Camargo, A., Braun, K., Slowik, R., Cavalli, I., Ribeiro, E., de Pedrosa, F., de Souza, E., Costa, F. and Klassen, G. (2010) 'Simultaneous CXCL12 and ESR1 CpG island hypermethylation correlates with poor prognosis in sporadic breast cancer', *BMC cancer,* 10, (1), pp. 23.

Ravelli, A. C. J., van der Meulen, J. H. P., Michels, R., Osmond, C., Barker, D. J. P., Hales, C. and Bleker, O. P. (1998) 'Glucose tolerance in adults after prenatal exposure to famine', *The Lancet,* 351, (9097), pp. 173-177.

Ravelli, A. C. J., van der Meulen, J. H. P., Osmond, C., Barker, D. J. P. and Bleker, O. P. (1999) 'Obesity at the age of 50 y in men and women exposed to famine prenatally', *The American Journal of Clinical Nutrition,* 70, (5), pp. 811.

Ravelli, G. P., Stein, Z. A. and Susser, M. W. (1976) 'Obesity in young men after famine exposure *in utero* and early infancy', *New England Journal of Medicine,* 295, (7), pp. 349-353.

Reik, W. (2007) 'Stability and flexibility of epigenetic gene regulation in mammalian development', *Nature,* 447, (7143), pp. 425-432.

Reitman, M. L., Mason, M. M., Moitra, J., Gavrilova, O., Marcus-Samuels, B., Eckhaus, M. and Vinson, C. (1999) 'Transgenic mice lacking white fat: models for understanding human lipoatrophic diabetes', *Annals of the New York Academy of Sciences,* 892, pp. 289-96.

Relton, C. L., Groom, A., Pourcain, B. S., Sayers, A. E., Swan, D. C., Embleton, N. D., Pearce, M. S., Ring, S. M., Northstone, K. and Tobias, J. H. (2012) 'DNA Methylation Patterns in Cord Blood DNA and Body Size in Childhood', *PloS one,* 7, (3), pp. e31821.

Relton, C. L. and Smith, G. D. (2010) 'Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment', *PLoS medicine,* 7, (10), pp. e1000356.

Relton, C. L. and Smith, G. D. (2012) 'Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease', *International journal of epidemiology,* 41, (1), pp. 161-176.

Remacle, C., Bieswal, F., Bol, V. and Reusens, B. (2011) 'Developmental programming of adult obesity and cardiovascular disease in rodents by maternal nutrition imbalance', *The American Journal of Clinical Nutrition,* 94, (6 Suppl), pp. 1846S-1852S.

Riant, E., Waget, A., Cogo, H., Arnal, J. F., Burcelin, R. and Gourdy, P. (2009) 'Estrogens protect against high-fat diet-induced insulin resistance and glucose intolerance in mice', *Endocrinology,* 150, (5), pp. 2109.

Richard, D., Guesdon, B. and Timofeeva, E. (2009) 'The brain endocannabinoid system in the regulation of energy balance', *Best Practice & Research Clinical Endocrinology & Metabolism,* 23, (1), pp. 17-32.

Richardson, B. (2003) 'Impact of aging on DNA methylation', *Ageing research reviews,* 2, (3), pp. 245-261.

Richon, V. M., Sandhoff, T. W., Rifkind, R. A. and Marks, P. A. (2000) 'Histone deacetylase inhibitor selectively induces p21WAF1 expression and gene-associated histone acetylation', *Proceedings of the National Academy of Sciences,* 97, (18), pp. 10014.

Rinaudo, P. (2012) 'Fetal Programming and Metabolic Syndrome', *Annual Review of Physiology,* 74, (1).

Ristow, M., Mulder, H., Pomplun, D., Schulz, T. J., Muller-Schmehl, K., Krause, A., Fex, M., Puccio, H., Muller, J. and Isken, F. (2003) 'Frataxin deficiency in pancreatic islets causes diabetes due to loss of beta cell mass', *Journal of Clinical Investigation,* 112, (4), pp. 527-534.

Ritov, V. B., Menshikova, E. V., Azuma, K., Wood, R., Toledo, F. G. S., Goodpaster, B. H., Ruderman, N. B. and Kelley, D. E. (2010) 'Deficiency of electron transport chain in human skeletal muscle mitochondria in type 2 diabetes mellitus and obesity', *American Journal of Physiology-Endocrinology And Metabolism,* 298, (1), pp. E49-E58.

Robert, M. F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I. C., Barsalou, A. and MacLeod, A. R. (2002) 'DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells', *Nature Genetics,* 33, (1), pp. 61-65.

Rodriguez, S., Gaunt, T. R., O'Dell, S. D., Chen, X., Gu, D., Hawe, E., Miller, G. J., Humphries, S. E. and Day, I. N. M. (2004) 'Haplotypic analyses of the IGF2-INS-TH gene cluster in relation to cardiovascular risk traits', *Human molecular genetics,* 13, (7), pp. 715.

Rodriguez-Acebes, S., Palacios, N., Botella-Carretero, J. I., Olea, N., Crespo, L., Peromingo, R., Gomez-Coronado, D., Lasuncian, M. A., Vazquez, C. and Martinez-Botas, J. (2010) 'Gene expression profiling of subcutaneous adipose tissue in morbid obesity using a focused microarray: Distinct expression of cell-cycle-and differentiation-related genes', *BMC medical genomics,* 3, (1), pp. 61.

Rogge, G., Jones, D., Hubert, G., Lin, Y. and Kuhar, M. (2008) 'CART peptides: regulators of body weight, reward and other functions', *Nature reviews. Neuroscience,* 9, (10), pp. 747.

Ronn, T., Poulsen, P., Hansson, O., Holmkvist, J., Almgren, P., Nilsson, P., Tuomi, T., Isomaa, B., Groop, L. and Vaag, A. (2008) 'Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle', *Diabetologia,* 51, (7), pp. 1159-1168.

Rosen, E. D., Sarraf, P., Troy, A. E., Bradwin, G., Moore, K., Milstone, D. S., Spiegelman, B. M. and Mortensen, R. M. (1999) 'PPAR-γ is required for the differentiation of adipose tissue in vivo and in vitro', *Molecular Cell,* 4, (4), pp. 611-617.

Rosenstock, J., Klaff, L. J., Schwartz, S., Northrup, J., Holcombe, J. H., Wilhelm, K. and Trautmann, M. (2010) 'Effects of exenatide and lifestyle modification on body weight and glucose tolerance in obese subjects with and without pre-diabetes', *Diabetes Care,* 33, (6), pp. 1173-1175.

Ross, R., Dagnone, D., Jones, P. J. H., Smith, H., Paddags, A., Hudson, R. and Janssen, I. (2000) 'Reduction in obesity and related comorbid conditions after diet-induced weight loss or exercise-induced weight loss in men', *Annals of Internal Medicine,* 133, (2), pp. 92-103.

Ruge, T., Hodson, L., Cheeseman, J., Dennis, A. L., Fielding, B. A., Humphreys, S. M., Frayn, K. N. and Karpe, F. (2009) 'Fasted to fed trafficking of fatty acids in human adipose tissue reveals a novel regulatory step for enhanced fat storage', *Journal of Clinical Endocrinology & Metabolism,* 94, (5), pp. 1781.

Sainsbury, A. and Zhang, L. (2010) 'Role of the arcuate nucleus of the hypothalamus in regulation of body weight during energy deficit', *Molecular and cellular endocrinology,* 316, (2), pp. 109-119.

Salomonis, N., Hanspers, K., Zambon, A., Vranizan, K., Lawlor, S., Dahlquist, K., Doniger, S., Stuart, J., Conklin, B. and Pico, A. (2007) 'GenMAPP 2: new features and resources for pathway analysis', *BMC bioinformatics,* 8, (1), pp. 217.

Samuelsson, A. M., Matthews, P. A., Argenton, M., Christie, M. R., McConnell, J. M., Jansen, E. H. J. M., Piersma, A. H., Ozanne, S. E., Twinn, D. F. and Remacle, C. (2008) 'Diet-induced obesity in female mice leads to offspring hyperphagia, adiposity, hypertension, and insulin resistance', *Hypertension,* 51, (2), pp. 383-392.

Sandovici, I., Leppert, M., Hawk, P. R., Suarez, A., Linares, Y. and Sapienza, C. (2003) 'Familial aggregation of abnormal methylation of parental alleles at the IGF2/H19 and IGF2R differentially methylated regions', *Human molecular genetics,* 12, (13), pp. 1569.

Santos, F., Hendrich, B., Reik, W. and Dean, W. (2002) 'Dynamic reprogramming of DNA methylation in the early mouse embryo', *Developmental biology,* 241, (1), pp. 172-182.

Santry, H. P., Gillen, D. L. and Lauderdale, D. S. (2005) 'Trends in bariatric surgical procedures', *JAMA: The Journal of the American Medical Association,* 294, (15), pp. 1909.

Sarwar, N., Aspelund, T., Eiriksdottir, G., Gobin, R., Seshasai, S. R. K., Forouhi, N. G., Sigurdsson, G., Danesh, J. and Gudnason, V. (2010) 'Markers of dysglycaemia and risk of

coronary heart disease in people without diabetes: Reykjavik prospective study and systematic review', *PLoS Medicine*, 7, (5), pp. e1000278.

Sassi, F. (2010) Obesity and the economics of prevention: fit not fat. *Publications de l'OCDE*.

Sato, N., Yamakawa, N., Masuda, M., Sudo, K., Hatada, I. and Muramatsu, M. (2011) 'Genome-wide DNA methylation analysis reveals phytoestrogen modification of promoter methylation patterns during embryonic stem cell differentiation', *PLoS ONE*, 6, (4), pp. e19278.

Scherer, T. and Buettner, C. (2011) 'Yin and Yang of hypothalamic insulin and leptin signaling in regulating white adipose tissue metabolism', *Reviews in Endocrine & Metabolic Disorders*, pp. 1-9.

Schernhammer, E. S., Giovannucci, E., Kawasaki, T., Rosner, B., Fuchs, C. S. and Ogino, S. (2010) 'Dietary folate, alcohol and B vitamins in relation to LINE-1 hypomethylation in colon cancer', *Gut,* 59, (19828464), pp. 794-799.

Schlesselman, J. J. and Stolley, P. D. (1982) Case-control studies: design, conduct, analysis. *Oxford University Press*, USA.

Schwartz, M. W., Woods, S. C., Porte, D., Seeley, R. J. and Baskin, D. G. (2000) 'Central nervous system control of food intake', *NATURE-LONDON-*, pp. 661-671.

Scouten, C. W. (2010) 'Frozen Section Technique in the Animal Research Setting', *A Practical Guide to Frozen Section Technique*, pp. 171-189.

Scuteri, A., Sanna, S., Chen, W. M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orru, M. and Usala, G. (2007) 'Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits', *PLoS genetics,* 3, (7), pp. e115.

Sears, R. A., Powers, L. S., Gerke, A., Philibert, R. and Monick, M. M. (2011) 'Smoking Altered DNA Methylation In Alveolar Macrophages Correlates With Gene Expression', *American Journal of Respiratory and Critical Care Medicine,* 183, (1 MeetingAbstracts), pp. A6446.

Sebert, S., Sharkey, D., Budge, H. and Symonds, M. E. (2011) 'The early programming of metabolic health: is epigenetic setting the missing link?', *American Journal of Clinical Nutrition*, 94, (6 Suppl), pp. 1953S-1958S.

Sedgwick, P. (2010) 'Nested case-control studies', *BMJ*, 340, pp. c2582.

Sengupta, P. K., Smith, E. M., Kim, K., Murnane, M. J. and Smith, B. D. (2003) 'DNA hypermethylation near the transcription start site of collagen 2 (I) gene occurs in both cancer cell lines and primary colorectal cancers', *Cancer research,* 63, (8), pp. 1789.

Shahkhalili, Y., Moulin, J., Zbinden, I., Aprikian, O. and MacÈ, K. (2010) 'Comparison of two models of intrauterine growth restriction for early catch-up growth and later development of glucose intolerance and obesity in rats', *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* 298, (1), pp. R141.

Sharma, P., Kumar, J., Garg, G., Kumar, A., Patowary, A., Karthikeyan, G., Ramakrishnan, L., Brahmachari, V. and Sengupta, S. (2008) 'Detection of altered global DNA methylation in coronary artery disease patients', *DNA Cell Biol,* 27, (18613790), pp. 357-365.

Shaw, K., Gennat, H., O'Rourke, P. and Del Mar, C. (2006) 'Exercise for overweight or obesity', *Cochrane Database of Systematic Reviews*, (4), pp. CD003817.

Shen, L., Guo, Y., Chen, X., Ahmed, S. and Issa, J. (2007a) 'Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis', *Biotechniques,* 42, (1), pp. 48.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R. A. and Issa, J. P. J. (2007b) 'Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters', *PLoS genetics,* 3, (10), pp. e181.

Simmons, D. (2011) 'Diabetes and obesity in pregnancy', *Best Practice & Research Clinical Obstetrics & Gynaecology,* 25, (1), pp. 25-36.

Sinclair, K. D., Allegrucci, C., Singh, R., Gardner, D. S., Sebastian, S., Bispham, J., Thurston, A., Huntley, J. F., Rees, W. D. and Maloney, C. A. (2007) 'DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status', *Proceedings of the National Academy of Sciences,* 104, (49), pp. 19351.

Singer, T., McConnell, M. J., Marchetto, M. C. N., Coufal, N. G. and Gage, F. H. (2010) 'LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes?', *Trends in neurosciences,* 33, (8), pp. 345-354.

Singhal, A., Kennedy, K., Lanigan, J., Fewtrell, M., Cole, T. J., Stephenson, T., Elias-Jones, A., Weaver, L. T., Ibhanesebhor, S. and MacDonald, P. D. (2010) 'Nutrition in infancy and long-term risk of obesity: evidence from 2 randomised controlled trials', *The American Journal of Clinical Nutrition,* 92, (5), pp. 1133.

Sinha, R., Fisch, G., Teague, B., Tamborlane, W. V., Banyas, B., Allen, K., Savoye, M., Rieger, V., Taksali, S. and Barbetta, G. (2002) 'Prevalence of impaired glucose tolerance among children and adolescents with marked obesity', *New England Journal of Medicine,* 346, (11), pp. 802-810.

Smirnov, D. A., Morley, M., Shin, E., Spielman, R. S. and Cheung, V. G. (2009) 'Genetic analysis of radiation-induced changes in human gene expression', *Nature,* 459, (7246), pp. 587.

Spalding, K. L., Arner, E., Westermark, P. O., Bernard, S., Buchholz, B. A., Bergmann, O., Blomqvist, L., Hoffstedt, J., N‰slund, E. and Britton, T. (2008) 'Dynamics of fat cell turnover in humans', *Nature,* 453, (7196), pp. 783-787.

Spencer, V. A. and Davie, J. R. (1999) 'Role of covalent modifications of histones in regulating gene expression', *Gene,* 240, (1), pp. 1-12.

Spiegelman, B. M. and Flier, J. S. (2001) 'Obesity and the regulation of energy balance', *Cell,* 104, pp. 531-543.

Stacey, S. N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S. A., Masson, G., Jakobsdottir, M., Thorlacius, S. and Helgason, A. (2007) 'Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptorñpositive breast cancer', *Nature genetics,* 39, (7), pp. 865-869.

Steegers-Theunissen, R. P., Obermann-Borst, S. A., Kremer, D., Lindemans, J., Siebel, C., Steegers, E. A., Slagboom, P. E. and Heijmans, B. T. (2009) 'Periconceptional maternal folic acid use of 400 μg per day is related to increased methylation of the IGF2 gene in the very young child', *PLoS One,* 4, (11), pp. e7845.

Stein, Z., Susser, M., Saenger, G. and Marolla, F. (1975) 'Famine and human development: The Dutch hunger winter of 1944-1945'. *Ann Intern Med* 83, (2) pp. 290

Stettler, N. and Iotova, V. (2010) 'Early growth patterns and long-term obesity risk', *Current Opinion in Clinical Nutrition & Metabolic Care,* 13, (3), pp. 294.

Stevens, A., Begum, G., Cook, A., Connor, K., Rumball, C., Oliver, M., Challis, J., Bloomfield, F. and White, A. (2010) 'Epigenetic changes in the hypothalamic

proopiomelanocortin and glucocorticoid receptor genes in the ovine fetus after periconceptional undernutrition', Endocrinology, 151, (8), pp. 3652-64.

Stoger, R., Kubicka, P., Liu, C. G., Kafri, T., Razin, A., Cedar, H. and Barlow, D. (1993) 'Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal', *Cell,* 73, (1), pp. 61-71.

Strahl, B. D., Briggs, S. D., Brame, C. J., Caldwell, J. A., Koh, S. S., Ma, H., Cook, R. G., Shabanowitz, J., Hunt, D. F. and Stallcup, M. R. (2001) 'Methylation of histone H4 at arginine three occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1', *Current Biology,* 11, (12), pp. 996-1000.

Suga, Y., Miyajima, K., Oikawa, T., Maeda, J., Usuda, J., KAJIWARA, N., OHIRA, T., UCHIDA, O., TSUBOI, M. and HIRANO, T. (2008) 'Quantitative p16 and ESR1 methylation in the peripheral blood of patients with non-small cell lung cancer', *Oncology reports,* 20, (5), pp. 1137-1142.

Sugawara, H., Iwamoto, K., Bundo, M., Ueda, J., Ishigooka, J. and Kato, T. (2011) 'Comprehensive DNA methylation analysis of human peripheral blood leukocytes and lymphoblastoid cell lines', *Epigenetics: official journal of the DNA Methylation Society,* 6, (4).

Swinburn, B. A., Sacks, G., Hall, K. D., McPherson, K., Finegood, D. T., Moodie, M. L. and Gortmaker, S. L. (2011) 'The global obesity pandemic: shaped by global drivers and local environments', *The Lancet,* 378, (9793), pp. 804-814.

Symonds, M. E. (2010) 'Nutrition and its contribution to obesity and diabetes: a life-course approach to disease prevention?', *Proceedings of the Nutrition Society,* 68, (1), pp. 71.

Symonds, M. E., Stephenson, T., Gardner, D. S. and Budge, H. (2006) 'Long-term effects of nutritional programming of the embryo and fetus: mechanisms and critical windows', *Reproduction, Fertility and Development,* 19, (1), pp. 53-63.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R. and Aravind, L. (2009) 'Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1', *Science,* 324, (5929), pp. 930.

Talens, R. P., Jukema, J., Trompet, S., Kremer, D., Westendorp, R., Lumey, L., Sattar, N., Putter, H., Slagboom, P. and Heijmans, B. (2011) 'Hypermethylation at loci sensitive to the prenatal environment is associated with increased incidence of myocardial infarction', *International journal of epidemiology* (41), pp. 106–115

Taveras, E. M., Rifas-Shiman, S. L., Belfort, M. B., Kleinman, K. P., Oken, E. and Gillman, M. W. (2009) 'Weight status in the first six months of life and obesity at three years of age', *Pediatrics,* 123, (4), pp. 1177.

Terry, M. B., Ferris, J. S., Pilsner, R., Flom, J. D., Tehranifar, P., Santella, R. M., Gamble, M. V. and Susser, E. (2008) 'Genomic DNA methylation among women in a multiethnic New York City birth cohort', *Cancer Epidemiology Biomarkers & Prevention,* 17, (9), pp. 2306-2310.

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Gayther, S. A., Apostolidou, S., Jones, A., Lechner, M., Beck, S., Jacobs, I. J. and Widschwendter, M. (2009) 'An epigenetic signature in peripheral blood predicts active ovarian cancer', *PLoS One,* 4, (12), pp. e8274.

Thompson, R. F., Atzmon, G., Gheorghe, C., Liang, H. Q., Lowes, C., Greally, J. M. and Barzilai, N. (2010a) 'Tissue specific dysregulation of DNA methylation in aging', *Aging cell,* 9, (4), pp. 506-518.

Thompson, R. F., Fazzari, M. J., Niu, H., Barzilai, N., Simmons, R. A. and Greally, J. M. (2010b) 'Experimental intrauterine growth restriction induces alterations in DNA methylation and gene expression in pancreatic islets of rats', *Journal of Biological Chemistry,* 285, (20), pp. 15111.

Tierling, S., Souren, N., Reither, S., Zang, K., Meng-Hentschel, J., Leitner, D., Oehl-Jaschkowitz, B. and Walter, J. (2011) 'DNA methylation studies on imprinted loci in a male monozygotic twin pair discordant for Beckwith-Wiedemann syndrome', *Clinical genetics,* 79, (6), pp. 546-553.

Tilling, K., Davies, N., Windmeijer, F., Kramer, M. S., Bogdanovich, N., Matush, L., Patel, R., Smith, G. D., Ben-Shlomo, Y. and Martin, R. M. (2011) 'Is infant weight associated with childhood blood pressure? Analysis of the Promotion of Breastfeeding Intervention Trial (PROBIT) cohort', *International journal of epidemiology,* 40, (5), pp. 1227-1237.

Tobi, E. W., Heijmans, B. T., Kremer, D., Putter, H., Delemarre-van de Waal, H. A., Finken, M. J. J., Wit, J. M. and Slagboom, P. E. (2011) 'DNA methylation of IGF2, GNASAS, INSIGF and LEP and being born small for gestational age', *Epigenetics,* 6, (2), pp. 171-176.

Tobi, E. W., Lumey, L., Talens, R. P., Kremer, D., Putter, H., Stein, A. D., Slagboom, P. E. and Heijmans, B. T. (2009) 'DNA methylation differences after exposure to prenatal famine are common and timing-and sex-specific', *Human molecular genetics,* 18, (21), pp. 4046.

Tominaga, K., Fujii, S., Mukawa, K., Fujita, M., Ichikawa, K., Tomita, S., Imai, Y., Kanke, K., Ono, Y. and Terano, A. (2005) 'Prediction of colorectal neoplasia by quantitative methylation analysis of estrogen receptor gene in nonneoplastic epithelium from patients with ulcerative colitis', *Clinical cancer research,* 11, (24), pp. 8880.

Tost, J. and Gut, I. G. (2007) 'DNA methylation analysis by pyrosequencing', *Nature protocols,* 2, (9), pp. 2265-2275.

Trayhurn, P. and Beattie, J. H. (2001) 'Physiological role of adipose tissue: white adipose tissue as an endocrine and secretory organ', *Proceedings of the Nutrition Society,* 60, (03), pp. 329-339.

Tsujikawa, M. and Tano, Y. (2007) 'M1S1 (TACSTD2) Development in the Vertebrate Genome', *Nippon Ganka Kiyo,* 58, (12), pp. 734-738.

Turan, N., Katari, S., Coutifaris, C. and Sapienza, C. (2010) 'Explaining inter-individual variability in phenotype: is epigenetics up to the challenge?', *Epigenetics: Official Journal of the DNA Methylation Society,* 5, (20083905), pp. 16-19.

Turcot, V., Groom, A., McConnell, J. C., Pearce, M. S., Potter, C., Embleton, N. D., Swan, D. C. and Relton, C. L. (2012) 'Bioinformatic selection of putative epigenetically regulated loci associated with obesity using gene expression data', *Gene*, 499, (1), pp. 99-107.

Ukkola, O., Sun, G. and Bouchard, C. (2001) 'Insulin-like growth factor 2 (IGF2) and IGF-binding protein 1 (IGFBP1) gene variants are associated with overfeeding-induced metabolic changes', *Diabetologia,* 44, (12), pp. 2231-2236.

Uribe-Lewis, S., Woodfine, K., Stojic, L. and Murrell, A. (2011) 'Molecular mechanisms of genomic imprinting and clinical implications for cancer', *Expert Reviews in Molecular Medicine,* 13, (1).

Utter, A. C., Nieman, D., Shannonhouse, E., Butterworth, D. and Nieman, C. (1998) 'Influence of diet and/or exercise on body composition and cardiorespiratory fitness in obese women', *International journal of sport nutrition,* 8, (3), pp. 213.

Vaissiere, T., Sawan, C. and Herceg, Z. (2008) 'Epigenetic interplay between histone modifications and DNA methylation in gene silencing', *Mutation Research,* 659, (1-2), pp. 40-8.

Van Den Bergh, B. E. A. R. H. (2011) 'Developmental programming of early brain and behaviour development and mental health: a conceptual framework', *Developmental Medicine & Child Neurology,* 53, pp. 19-23.

van Dijk, S. J., Feskens, E. J. M., Bos, M. B., Hoelen, D. W. M., Heijligenberg, R., Bromhaar, M. G., de Groot, L. C., de Vries, J. H. M., M√°ller, M. and Afman, L. A. (2009) 'A saturated fatty acid-rich diet induces an obesity-linked proinflammatory gene expression profile in adipose tissue of subjects at risk of metabolic syndrome', *The American Journal of Clinical Nutrition,* 90, (6), pp. 1656-1664.

van Iersel, M., Kelder, T., Pico, A., Hanspers, K., Coort, S., Conklin, B. and Evelo, C. (2008) 'Presenting and exploring biological pathways with PathVisio', *BMC bioinformatics,* 9, (1), pp. 399.

Varley, K. E. and Mitra, R. D. (2010) 'Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples', *Genome research,* 20, (9), pp. 1279.

Vasiljevic, N., Wu, K., Brentnall, A. R., Kim, D. C., Thorat, M. A., Kudahetti, S. C., Mao, X., Xue, L., Yu, Y. and Shaw, G. L. (2011) 'Absolute quantitation of DNA methylation of 28 candidate genes in prostate cancer using pyrosequencing', *Disease Markers,* 30, (4), pp. 151-161.

Vehaskari, V. M. (2010) 'Prenatal programming of kidney disease', *Current opinion in pediatrics,* 22, (2), pp. 176.

Vickers, M., Gluckman, P., Coveny, A., Hofman, P., Cutfield, W., Gertler, A., Breier, B. and Harris, M. (2005) 'Neonatal leptin treatment reverses developmental programming', *Endocrinology,* 146, (10), pp. 4211-4216.

Vickers, M. H. (2011) 'Developmental programming of the metabolic syndrome-critical windows for intervention', *World Journal of Diabetes,* 2, (9), pp. 137.

Vickers, M. H. and Sloboda, D. M. (2010) 'Prenatal nutritional influences on obesity risk in offspring', *Nutrition and Dietary Supplements,* 2, pp. 137-149.

Visscher, T. L. S., Snijder, M. B. and Seidell, J. C. (2010) 'Epidemiology: Definition and Classification of Obesity', *Clinical Obesity in Adults and Children*, pp. 1-14.

Von Kries, R., Koletzko, B., Sauerwald, T., Von Mutius, E., Barnert, D., Grunert, V. and Von Voss, H. (1999) 'Breast feeding and obesity: cross sectional study', *Bmj,* 319, (7203), pp. 147.

Vong, L., Ye, C., Yang, Z., Choi, B., Chua Jr, S. and Lowell, B. B. (2011) 'Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons', *Neuron,* 71, (1), pp. 142-154.

Vucetic, Z., Kimmel, J., Totoki, K., Hollenbeck, E. and Reyes, T. M. (2010) 'Maternal high-fat diet alters methylation and gene expression of dopamine and opioid-related genes', *Endocrinology,* 151, (10), pp. 4756.

Wadsworth, M. (2002) 'Birth cohort studies', *Encyclopedia of Biostatistics.*

Walsh, S. W. (2007) 'Obesity: a risk factor for preeclampsia', *Trends in Endocrinology & Metabolism,* 18, (10), pp. 365-370.

Wan, Y.-j., Li, Y.-y., Xia, W., Chen, J., Lv, Z.-q., Zeng, H.-c., Zhang, L., Yang, W.-j., Chen, T., Lin, Y., Wei, J. and Xu, S.-q. (2010) 'Alterations in tumor biomarker GSTP gene

methylation patterns induced by prenatal exposure to PFOS', *Toxicology,* 274, (20621739), pp. 57-64.

Wang, L., Pal, S. and Sif, S. (2008) 'Protein arginine methyltransferase 5 suppresses the transcription of the RB family of tumor suppressors in leukemia and lymphoma cells', *Molecular and cellular biology,* 28, (20), pp. 6262.

Wang, Q., Bing, C., Al-Barazanji, K., Mossakowaska, D. E., Wang, X. M., McBay, D., Neville, W., Taddayon, M., Pickavance, L. and Dryden, S. (1997) 'Interactions between leptin and hypothalamic neuropeptide Y neurons in the control of food intake and energy homeostasis in the rat', *Diabetes,* 46, (3), pp. 335-341.

Wang, R., Liu, X., Hentges, S., Dunn-Meynell, A., Levin, B., Wang, W. and Routh, V. (2004) 'The regulation of glucose-excited neurons in the hypothalamic arcuate nucleus by glucose and feeding-relevant peptides', *Diabetes,* 53, (8), pp. 1959.

Wang, X., Zhu, H., Snieder, H., Su, S., Munn, D., Harshfield, G., Maria, B., Dong, Y., Treiber, F. and Gutin, B. (2010a) 'Obesity related methylation changes in DNA of peripheral blood leukocytes', *BMC medicine,* 8, (1), pp. 87.

Wang, Y. (2004) 'Evolution mechanisms and biological significance of CG dinucleotide compositional bias in genomes', *HKU Theses Online (HKUTO).*

Wang, Y., Fu, J. L., Song, W. W. and Wang, L. L. (2010b) 'Genomic imprinting status of IGF-II and H19 in placentas of fetal growth restriction patients', *Journal of genetics,* 89, (2), pp. 213-216.

Waterland, R. A. (2005) 'Does nutrition during infancy and early childhood contribute to later obesity via metabolic imprinting of epigenetic gene regulatory mechanisms?'*, Nestle Nutrition Workshop Series. Paediatric Programme*, 56, pp. 157-71; discussion 171-4.

Waterland, R. A. (2009) 'Is epigenetics an important link between early life events and adult disease?', *Hormone Research in Paediatrics,* 71, (1), pp. 13-16.

Waterland, R. A. and Jirtle, R. L. (2003) 'Transposable elements: targets for early nutritional effects on epigenetic gene regulation', *Mol Cell Biol,* 23, (12861015), pp. 5293-5300.

Waterland, R. A., Travisano, M. and Tahiliani, K. G. (2007) 'Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female', *The FASEB Journal,* 21, (12), pp. 3380.

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., Dymov, S., Szyf, M. and Meaney, M. J. (2004) 'Epigenetic programming by maternal behavior', *Nature neuroscience,* 7, (8), pp. 847-854.

Wei, Y., Yu, L., Bowen, J., Gorovsky, M. A. and Allis, C. D. (1999) 'Phosphorylation of histone H3 is required for proper chromosome condensation and segregation', *Cell,* 97, (1), pp. 99-109.

Weisenberger, D. J., Campan, M., Long, T. I., Kim, M., Woods, C., Fiala, E., Ehrlich, M. and Laird, P. W. (2005) 'Analysis of repetitive element DNA methylation by MethyLight', *Nucleic Acids Research,* 33, (21), pp. 6823.

Whincup, P. H., Kaye, S. J., Owen, C. G., Huxley, R., Cook, D. G., Anazawa, S., Barrett-Connor, E., Bhargava, S. K., Birgisdottir, B. E. and Carlsson, S. (2008) 'Birth weight and risk of Type 2 diabetes', *JAMA: The Journal of the American Medical Association,* 300, (24), pp. 2886.

Whitelaw, N. C. and Whitelaw, E. (2008) 'Transgenerational epigenetic inheritance in health and disease', *Current opinion in genetics & development,* 18, (3), pp. 273-279.

Great Britain. Parliament. Organisation, W. H., (1990) *Monitoring trends and determinants in cardiovascular disease project.* World Health Organisation.

WHO (2011) *Obesity and overweight - Factsheet.* Available at: http://www.who.int/mediacentre/factsheets/fs311/en/ (Accessed: 11/01/12).

Widiker, S., K√§rst, S., Wagener, A. and Brockmann, G. (2010) 'High-fat diet leads to a decreased methylation of theMc4r gene in the obese BFMI and the lean B6 mouse lines', *Journal of applied genetics,* 51, (2), pp. 193-197.

Widschwendter, M., Siegmund, K. D., M¸ller, H. M., Fiegl, H., Marth, C., M¸ller-Holzner, E., Jones, P. A. and Laird, P. W. (2004) 'Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen', *Cancer research,* 64, (11), pp. 3807.

Wierup, N., Richards, W., Bannon, A., Kuhar, M., AhrÈn, B. and Sundler, F. (2005) 'CART knock out mice have impaired insulin secretion and glucose intolerance, altered beta cell morphology and increased body weight', *Regulatory peptides,* 129, (1-3), pp. 203-211.

Williams, G., Bing, C., Cai, X. J., Harrold, J. A., King, P. J. and Liu, X. H. (2001) 'The hypothalamus and the control of energy homeostasis:: Different circuits, different purposes', *Physiology & behavior,* 74, (4-5), pp. 683-701.

Williams, J. (1981) 'Catch-up growth', *Journal of Embryology and Experimental Morphology,* 65, (Supplement), pp. 89-101.

Williams, K. T. and Schalinske, K. L. (2012) 'Tissue-specific alterations of methyl group metabolism with DNA hypermethylation in the Zucker (type 2) diabetic fatty rat', *Diabetes/Metabolism Research Reviews*, 28, (2), pp. 123-31.

Wilson, B. D., Ollmann, M. M. and Barsh, G. S. (1999) 'The role of agouti-related protein in regulating body weight', *Molecular medicine today,* 5, (6), pp. 250-256.

Wolffe, A. (1998) *Chromatin: structure and function.* Academic Press San Diego, CA:.

Wong, C., Caspi, A., Williams, B., Craig, I. W., Houts, R., Ambler, A., Moffitt, T. E. and Mill, J. (2010) 'A longitudinal study of epigenetic variation in twins', *Epigenetics: official journal of the DNA Methylation Society,* 5, (6), pp. 516.

Wood, I. S., de Heredia, F. P., Wang, B. and Trayhurn, P. (2009) 'Cellular hypoxia and adipose tissue dysfunction in obesity', *Proc Nutr Soc,* 68, (4), pp. 370-7.

Woodson, K., Flood, A., Green, L., Tangrea, J. A., Hanson, J., Cash, B., Schatzkin, A. and Schoenfeld, P. (2004) 'Loss of insulin-like growth factor-II imprinting and the presence of screen-detected colorectal adenomas in women', *Journal of the National Cancer Institute,* 96, (5), pp. 407.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W. and Walter, J. (2011) '5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming', *Nature Communications,* 2, pp. 241.

Wright, C., Avery, A., Epstein, M., Birks, E. and Croft, D. (1998) 'New chart to evaluate weight faltering', *Archives of disease in childhood,* 78, (1), pp. 40.

Wright, C., Matthews, J., Waterston, A. and Aynsley Green, A. (1994) 'What is a normal rate of weight gain in infancy?', *Acta Paediatrica,* 83, (4), pp. 351-356.

Wright, R. O., Schwartz, J., Wright, R. J., Bollati, V., Tarantini, L., Park, S. K., Hu, H., Sparrow, D., Vokonas, P. and Baccarelli, A. (2010) 'Biomarkers of lead exposure and DNA methylation within retrotransposons', *Environmental Health Perspectives,* 118, (6), pp. 790.

Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E. and Zhang, Y. (2011a) 'Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells', *Genes & development,* 25, (7), pp. 679.

Wu, H. C., Delgado-Cruzata, L., Flom, J. D., Kappil, M., Ferris, J. S., Liao, Y., Santella, R. M., Terry, M. B. and Wu, H. (2011b) 'Global methylation profiles in DNA from different blood cell types', *Epigenetics,* 6, (1), pp. 76.

Wylie, A. A., Pulford, D. J., McVie-Wylie, A. J., Waterland, R. A., Evans, H. K., Chen, Y. T., Nolan, C. M., Orton, T. C. and Jirtle, R. L. (2003) 'Tissue-specific inactivation of murine M6P/IGF2R', *The American journal of pathology,* 162, (1), pp. 321.

Xiao, Y., Word, B., Starlard-Davenport, A., Haefele, A., Lyn-Cook, B. D. and Hammons, G. (2008) 'Age and gender affect DNMT3a and DNMT3b expression in human liver', *Cell biology and toxicology,* 24, (3), pp. 265-272.

Xu, J., Pope, S. D., Jazirehi, A. R., Attema, J. L., Papathanasiou, P., Watts, J. A., Zaret, K. S., Weissman, I. L. and Smale, S. T. (2007) 'Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells', *Proceedings of the National Academy of Sciences,* 104, (30), pp. 12377.

Yajnik, C. S. and Deshmukh, U. S. (2008) 'Maternal nutrition, intrauterine programming and consequential risks in the offspring', *Reviews in endocrine & metabolic disorders,* 9, (3), pp. 203-211.

Yang, A. S., Estacio, M. R. H., Doshi, K., Kondo, Y., Tajara, E. H. and Issa, J. P. J. (2004) 'A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements', *Nucleic Acids Research,* 32, (3), pp. e38-e38.

Yang, B., Sun, H., Lin, W., Hou, W., Li, H., Zhang, L., Li, F., Gu, Y., Song, Y., Li, Q. and Zhang, F. (2011) 'Evaluation of global DNA hypomethylation in human prostate cancer and prostatic intraepithelial neoplasm tissues by immunohistochemistry', *Urol Oncol*, 23, pp. 23.

Yi, J. M., Dhir, M., Guzzetta, A. A., Iacobuzio-Donahue, C. A., Heo, K., Yang, K. M., Suzuki, H., Toyota, M., Kim, H. M. and Ahuja, N. (2012) 'DNA methylation biomarker candidates for early detection of colon cancer', *Tumor Biology*, pp. 1-10.

Yoo, S. B., Ryu, V., Park, E. Y., Kim, B. T., Kang, D. W., Lee, J. H. and Jahng, J. W. (2011) 'The arcuate NPY, POMC, and CART expressions responding to food deprivation are exaggerated in young female rats that experienced neonatal maternal separation', *Neuropeptides*, 45, (5), pp. 343-9.

Young, L., Fernandes, K. and McEvoy, T. (2001) 'Epigenetic change in IGF2R is associated with fetal overgrowth after sheep embryo culture', *Nature genetics,* 27, pp. 153.

Yu, Z. B., Han, S. P., Zhu, G. Z., Zhu, C., Wang, X. J., Cao, X. G. and Guo, X. R. (2011) 'Birth weight and subsequent risk of obesity: a systematic review and meta-analysis', *Obesity Reviews*, 12, (7), pp. 525-42.

Yura, S., Itoh, H., Sagawa, N., Yamamoto, H., Masuzaki, H., Nakao, K., Kawamura, M., Takemura, M., Kakui, K. and Ogawa, Y. (2005) 'Role of premature leptin surge in obesity resulting from intrauterine undernutrition', *Cell metabolism,* 1, (6), pp. 371-378.

Zeng, H., Yan, L., Cheng, W. H. and Uthus, E. O. (2011) 'Dietary Selenomethionine Increases Exon-Specific DNA Methylation of the p53 Gene in Rat Liver and Colon Mucosa', *The Journal of nutrition,* 141, (8), pp. 1464.

Zhang, B. and Yao, Y. F. (2010) 'Gelatinous drop-like corneal dystrophy with a novel mutation of TACSTD2 manifested in combination with spheroidal degeneration in a Chinese patient', *Molecular vision,* 16, pp. 1570.

Zhang, D., Cheng, L., Badner, J. A., Chen, C., Chen, Q., Luo, W., Craig, D. W., Redman, M., Gershon, E. S. and Liu, C. (2010a) 'Genetic control of individual differences in gene-specific methylation in human brain', *The American Journal of Human Genetics,* 86, (3), pp. 411-419.

Zhang, F. F., Cardarelli, R., Carroll, J., Zhang, S., Fulda, K. G., Gonzalez, K., Vishwanatha, J. K., Morabia, A. and Santella, R. M. (2011a) 'Physical activity and global genomic DNA methylation in a cancer-free population', *Epigenetics: Official Journal of the DNA Methylation Society,* 6, (3), pp. 293-9.

Zhang, F. F., Morabia, A., Carroll, J., Gonzalez, K., Fulda, K., Kaur, M., Vishwanatha, J. K., Santella, R. M. and Cardarelli, R. (2011b) 'Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population', *Journal of Nutrition,* 141, (6), pp. 1165-71.

Zhang, S., Rattanatray, L., McMillen, I. C., Suter, C. M. and Morrison, J. L. (2011) 'Periconceptional nutrition and the early programming of a life of obesity or adversity', *Progress in Biophysics & Molecular Biology*, 106, (1), pp. 307-14.

Zhang, X., Zhang, G., Zhang, H., Karin, M., Bai, H. and Cai, D. (2008) 'Hypothalamic IKK [beta]/NF-[kappa] B and ER stress link overnutrition to energy imbalance and obesity', *Cell,* 135, (1), pp. 61-73.

Zhao, J., Goldberg, J., Bremner, J. D. and Vaccarino, V. (2012) 'Global DNA Methylation Is Associated With Insulin Resistance', *Diabetes*, 61, (2), pp. 542-546.

Zhong, H., Yang, X., Kaplan, L. M., Molony, C. and Schadt, E. E. (2010) 'Integrating pathway analysis and genetics of gene expression for genome-wide association studies', *The American Journal of Human Genetics,* 86, (4), pp. 581-591.

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. and Henikoff, S. (2006) 'Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription', *Nature Genetics,* 39, (1), pp. 61-69.

Zvi, A., Ariel, N., Fulkerson, J., Sadoff, J. C. and Shafferman, A. (2008) 'Whole genome identification of Mycobacterium tuberculosis vaccine candidates by comprehensive data mining and bioinformatic analyses', *BMC Medical Genomics,* 1, (1), pp. 18.

Appendix 1: Presentations of methods or results from this thesis

2011 North East postgraduate research conference, Newcastle

Using Ingenuity to identify differentially methylated candidates from an expression array dataset.

2010 European Nutrigenomics Organisation week, Glasgow

Global DNA methylation is associated with bloody lipid profiles at age 50.

2009 European Nutrigenomics Organisation week, Montecatini Terme, Italy

Lifecourse influences on DNA methylation.

2009 International Congress of Nutrition, Bangkok, Thailand

Epigenetic Epidemiology: Evidence for the Role of Epigenetic Variation in Complex Disease.

Appendix 2: Original research articles published using methods or results from this thesis

International Journal of Epidemiology Volume 41, Issue 1 Pp. 210-217.

Global LINE-1 DNA methylation is associated with blood glycaemic and lipid profiles

Mark S Pearce[1], James C McConnell[2], Catherine Potter[2], Laura M Barrett[2], Louise Parker [3,4], John C Mathers[5], and Caroline L Relton[2]

[1]Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK, [2]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK, [3]Department of Medicine, [4]Department of Paediatrics, Dalhousie University, Halifax, Canada, [5]Human Nutrition Research Centre, Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne, UK

Accepted January 30, 2012.

Abstract

Background Patterns of DNA methylation change with age and these changes are believed to be associated with the development of common complex diseases. The hypothesis that Long Interspersed Nucleotide Element 1 (LINE-1) DNA methylation (an index of global DNA methylation) is associated with biomarkers of metabolic health was investigated in this study.

Methods Global LINE-1 DNA methylation was quantified by pyrosequencing in blood-derived DNA samples from 228 individuals, aged 49–51 years, from the Newcastle Thousand Families Study (NTFS). Associations between log-transformed LINE-1 DNA methylation levels and anthropometric and blood biochemical measurements, including triglycerides, total cholesterol, low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol, fasting glucose and insulin secretion and resistance were examined.

Results Linear regression, after adjustment for sex, demonstrated positive associations between log-transformed LINE-1 DNA methylation and fasting glucose {coefficient 2.80 [95% confidence interval (CI) 0.39–5.22]}, total cholesterol [4.76 (95% CI 1.43–8.10)], triglycerides [3.83 (95% CI 1.30–6.37)] and LDL-cholesterol [5.38 (95% CI 2.12–8.64)] concentrations. A negative association was observed between log-transformed LINE-1 methylation and both HDL cholesterol concentration [−1.43 (95% CI −2.38 to −0.48)] and HDL:LDL ratio [−1.06 (95% CI −1.76 to −0.36)]. These coefficients reflect the millimoles per litre change in biochemical measurements per unit increase in log-transformed LINE-1 methylation.

Conclusions These novel associations between global LINE-1 DNA methylation and blood glycaemic and lipid profiles highlight a potential role for epigenetic biomarkers as predictors of metabolic disease and may be relevant to future diagnosis, prevention and treatment of this group of disorders. Further work is required to establish the role of confounding and reverse causation in the observed associations.

Gene Available online 8 March 2012

Bioinformatic selection of putative epigenetically regulated loci associated with obesity using gene expression data

Valérie Turcot[1], Alexandra Groom[1], James C. McConnell[1], Mark S. Pearce[2], Catherine Potter[1], Nicholas D. Embleton[3], Daniel C. Swan[4], Caroline L. Relton[1]

[1]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK, [2]Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK, [3]Neonatal Service, Royal Victoria Infirmary, Newcastle upon Tyne, UK, [4]Bioinformatics Unit, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, UK

Accepted February 1, 2012.

Abstract

There is considerable interest in defining the relationship between epigenetic variation and the risk of common complex diseases. Strategies which assist in the prioritisation of target loci that have the potential to be epigenetically regulated might provide a useful approach in identifying concrete examples of epigenotype–phenotype associations. Focusing on the postulated role of epigenetic factors in the aetiopathogenesis of obesity this report outlines an approach utilising gene expression data and a suite of bioinformatic tools to prioritise a list of target candidate genes for more detailed experimental scrutiny. Gene expression microarrays were performed using peripheral blood RNA from children aged 11–13 years selected from the Newcastle Preterm Birth Growth Study which were grouped by body mass index (BMI). Genes showing ≥ 2.0 fold differential expression between low and high BMI groups were selected for *in silico* analysis. Several bioinformatic tools were used for each following step; 1) a literature search was carried out to identify whether the differentially expressed genes were associated with adiposity phenotypes. Of those obesity-candidate genes, putative epigenetically regulated promoters were identified by 2) defining the promoter regions, 3) then by selecting promoters with a CpG island (CGI), 4) and then by identifying any transcription factor binding modules covering CpG sites within the CGI. This bioinformatic processing culminated in the identification of a short list of target obesity-candidate genes putatively regulated by DNA methylation which can be taken forward for experimental analysis. The proposed workflow provides a flexible, versatile and low cost methodology for target gene prioritisation that is applicable to multiple species and disease contexts.

Highlights

► We proposed *in silico* tools to target obesity candidate loci putatively regulated epigenetically. ► Microarrays revealed differentially expressed genes between variable adiposity index. ► Further selection of target loci was achieved using a gene-adiposity literature search. ► Promoters of target loci were analysed for potential methylation regulation

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology 2012;**41**:210–217
doi:10.1093/ije/dys020

Global LINE-1 DNA methylation is associated with blood glycaemic and lipid profiles

Mark S Pearce,[1]* James C McConnell,[2] Catherine Potter,[2] Laura M Barrett,[2] Louise Parker,[3,4] John C Mathers[5] and Caroline L Relton[2]

[1]Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK, [2]Human Nutrition Research Centre, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK, [3]Department of Medicine, [4]Department of Paediatrics, Dalhousie University, Halifax, Canada and [5]Human Nutrition Research Centre, Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne, UK

*Corresponding author. Institute of Health and Society, Newcastle University, Sir James Spence Institute, Royal Victoria Infirmary, Newcastle upon Tyne, NE1 4LP, UK. E-mail mark.pearce@ncl.ac.uk

Accepted 30 January 2012

Background Patterns of DNA methylation change with age and these changes are believed to be associated with the development of common complex diseases. The hypothesis that Long Interspersed Nucleotide Element 1 (LINE-1) DNA methylation (an index of global DNA methylation) is associated with biomarkers of metabolic health was investigated in this study.

Methods Global LINE-1 DNA methylation was quantified by pyrosequencing in blood-derived DNA samples from 228 individuals, aged 49–51 years, from the Newcastle Thousand Families Study (NTFS). Associations between log-transformed LINE-1 DNA methylation levels and anthropometric and blood biochemical measurements, including triglycerides, total cholesterol, low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol, fasting glucose and insulin secretion and resistance were examined.

Results Linear regression, after adjustment for sex, demonstrated positive associations between log-transformed LINE-1 DNA methylation and fasting glucose {coefficient 2.80 [95% confidence interval (CI) 0.39–5.22]}, total cholesterol [4.76 (95% CI 1.43–8.10)], triglycerides [3.83 (95% CI 1.30–6.37)] and LDL-cholesterol [5.38 (95% CI 2.12–8.64)] concentrations. A negative association was observed between log-transformed LINE-1 methylation and both HDL cholesterol concentration [−1.43 (95% CI −2.38 to −0.48)] and HDL:LDL ratio [−1.06 (95% CI −1.76 to −0.36)]. These coefficients reflect the millimoles per litre change in biochemical measurements per unit increase in log-transformed LINE-1 methylation.

Conclusions These novel associations between global LINE-1 DNA methylation and blood glycaemic and lipid profiles highlight a potential role for epigenetic biomarkers as predictors of metabolic disease and may be relevant to future diagnosis, prevention and treatment of this group of disorders. Further work is required to establish the role of confounding and reverse causation in the observed associations.

Keywords Global DNA methylation, LINE-1, cohort study, glucose, HDL/LDL cholesterol, insulin, triglyceride

Introduction

DNA methylation is the covalent modification of cytosine residues in the DNA sequence through the addition of a methyl group that converts cytosine to 5-methyl cytosine (5meC).[1] In the human genome, this covalent modification largely takes place on cytosine residues that are located 5′ adjacent to guanine residues. These sequences of nucleotides are known as CpG sites, and in some regions of the genome they cluster together forming motifs known as CpG islands (CGIs). These CGIs make up a maximum of 2% of the genome, and are for the most part unmethylated. Some CGIs, however, are more highly methylated and these tend to be proximal to imprinted genes or transposons.[2] At a global level, DNA methylation can be assessed by utilizing repeat interspersed regions such as *Alu* or Long Interspersed Nucleotide Element 1 (LINE-1). Comprising ~17% of the human genome, LINE-1 is the most abundant family of non-long terminal repeat retrotransposons found in the genome.[3] Such elements have served as a useful proxy for global DNA methylation as they are commonly heavily methylated in normal tissue (although hypomethylated in tumour tissue), and are spread ubiquitously throughout the genome.[4] The level of correlation of LINE-1 with gene-specific methylation is not well documented, although LINE-1 methylation does correlate with global methylation measured using a variety of different methods including *Alu*, *Sat2* and LUMA.[5] Global hypomethylation is a common event in ageing cells. This has been shown in relation to some interspersed repeat regions such as *Alu*, but the evidence is less clear with regard to LINE-1.[4] Indeed, a recent study of age-dependent changes in DNA methylation of interspersed repeat regions showed a weak positive correlation between age and LINE-1 methylation.[6]

Limited evidence exists to link LINE-1 methylation with disease, and this has almost exclusively been observed in the cancer field. For example, LINE-1 methylation levels modulate the effects of exposure to dietary folate and alcohol on colon cancer risk,[7] and have prognostic value when analysed in colon tumour tissue.[8,9] In a recent study of LINE-1 methylation and cardiovascular health in the Boston-based Normative Aging Study, people with prevalent ischaemic heart disease (IHD) and stroke had lower LINE-1 methylation and, in longitudinal analyses, those with lower LINE-1 methylation were at higher risk for incident IHD, stroke and total mortality.[10] Thus, there is little empirical evidence to date linking LINE-1 methylation with common complex diseases other than cancer and limited evidence for an association with neural tube defects.[11] However, studies both in experimental animals and in human cohorts have shown that environmental insults can influence DNA methylation (reviewed by Mathers *et al.*[12]). LINE-1 methylation is also susceptible to a wide range of environmental exposures including perfluorooctane sulfonate,[13] prenatal tobacco smoke exposure,[14] polycyclic aromatic hydrocarbons,[15] biomarkers of lead levels in both adults[16] and cord blood,[17] traffic particulates[18] and plasma homocysteine.[19]

The present study addresses the hypothesis that global LINE-1 DNA methylation, measured at age 49–51 years, is associated with traits indicative of early-stage metabolic disease.

Methods

Study participants

The Newcastle Thousand Families Study (NTFS) prospective birth cohort consists of all 1142 individuals born in May and June 1947 to mothers resident within the city of Newcastle upon Tyne in northern England.[20] Two-thirds of these children were followed up regularly until age 15 years, with detailed information collected prospectively on their health, growth and socio-economic circumstances. Follow-up was re-established during the 1990s with participants being traced via media publicity or through the UK National Health Service Central Register. Between October 1996 and December 1998, 412 participants (~50 years of age) attended clinical examinations which included blood collection for DNA analysis and completed questionnaires detailing their family history and lifestyle.[20] DNA from 228 individuals was analysed in the current study, based upon samples of sufficient quality and quantity for LINE-1 DNA methylation analysis. Excluding sex, these 228 individuals were representative of the initial 1142 participants as well as the 412 follow-up participants (data not shown).

Clinical assessments of outcomes and adult height and weight at age 49–51 years

Assessments were performed in the morning following an overnight fast. All lipid analyses were performed on a DAX analyser (Bayer, Basingstoke). Total cholesterol was measured using a cholesterol oxidase/peroxidise method with calibrants traceable to the Centres for Disease Control definitive method. Serum high-density lipoprotein (HDL) cholesterol was measured using a cholesterol oxidase method after precipitation of apolipoprotein B containing lipoproteins with phosphotungstic acid and magnesium chloride (interassay coefficient of variation 2.2%). Low-density lipoprotein (LDL) cholesterol levels were derived by the Friedewald method[21] and the HDL:LDL ratio was calculated. Triglyceride concentrations, excluding glycerol, were estimated by a lipase-glycerol kinase method. Plasma glucose concentrations at 0, 30 and 120 min (after a 75-g oral glucose load) were measured on a Yellow Springs Analyser (YSI Stat Plus 2300; Yellow Springs Instruments, Farnborough, UK)[22] and serum insulin at the same time-points

was determined by ELISA (Dako Ltd, Ely, UK) (interassay coefficients of variation 3.1 and 3.3%, respectively).[23] Insulin resistance was estimated using the homeostasis model assessment of insulin resistance (HOMA-IR) of Matthews *et al*.[24] Insulin secretion was estimated as the ratio of the 30-min increment in insulin concentration to the 30-min increment in glucose concentration following oral glucose loading, relative to the baseline concentrations.[25] Height and weight were measured and body mass index (BMI) was calculated. Waist and hip circumferences were measured according to the protocol of the World Health Organization Monitoring Trends and Determinants in Cardiovascular Disease (MONICA) project.[26] Percent body fat was estimated from impedance measured using a Holtain body composition analyser (Holtain Ltd, Crymych, Wales, UK).

Measurement of global LINE-1 DNA methylation

DNA was extracted from peripheral blood samples using a Nucleon BACC2 kit (Tepnel Life Sciences, UK). One microgram of DNA sample was bisulphite modified using the Zymo EZ DNA Methylation Gold kit (Cambridge Bioscience, Cambridge) using the manufacturer's standard protocol. One microgram of bisulphite modified DNA was PCR amplified using 2 × HotstarTAQ Mastermix (Qiagen, UK), 2 mM $MgCl_2$ (Qiagen, UK) and 0.2 μM of each primer (LINE-1 forward primer—5′-TTT TGA GTT AGG TGT GGG ATA TA-3′ and LINE-1 reverse primer—BIO-5′-AAA ATC AAA AAA TTC CCT TTC-3′).[27] PCR conditions were as follows: 95°C for 15 min, 50 cycles of 95°C for 15 s, 60°C for 30 s and 72°C for 15 s and finally 72°C for 5 min. Five microlitres of amplicons were utilized for downstream single-strand preparation and hybridization of 0.5 μM sequencing primer (5′-GGG TGG GAG TGA T-3′), using a vacuum prep tool and workstation according to manufacturer's instructions (Qiagen, UK). LINE-1 methylation was quantified using a PyroMark MD Pyrosequencer (Qiagen, UK) in which the analysis sequence for LINE-1 was: TC/TGATTTTTTAGGTGC/TGTTC/TGTTA. Zero and 100% methylated controls were generated by carrying out a nested PCR reaction on genomic DNA to generate an unmethylated control, followed by *in vitro* methylation (*SssI* treatment) of an aliquot of the PCR product to generate a methylated control. These controls were used to rule out any amplification bias of primers for methylated DNA and to assess assay reproducibility. LINE-1 primer sets were found to be unbiased and were reproducible. 0 and 100% methylated controls were routinely run alongside samples as internal controls. Samples were analysed in duplicate with appropriate quality control measures in place (bisulphite conversion and PCR controls and random repeats). The mean (standard deviation) difference between duplicate samples was 3.0 (3.4%). This assay arbitrarily amplifies LINE-1 sequences from multiple genomic locations, providing a representative measure of methylation that is not site-specific i.e. a global assessment.

Statistical analysis

Pyrosequencing generated estimates of LINE-1 DNA methylation at each of three CpG sites which were expressed as a percentage, i.e. the proportion of methylated residues in the total DNA sample assayed. Correlation between methylation at all three CpG sites was high ($P < 0.001$), therefore a mean of all three sites was calculated and tested for association with the anthropometric, glycaemic and lipid-related outcome variables. Mean methylation levels were comparable with those previously published using a similar (but not identical) assay design.[28] DNA methylation showed a skewed distribution (tested using a ShapiroWilk test) so values were log-transformed before further analysis. Linear regression was used to examine relationships between anthropometric measures, glycaemic and lipid-related blood biomarkers (the dependent variables) and the independent variable, log-transformed global LINE-1 DNA methylation at age 50 years. Regression coefficients and corresponding 95% confidence intervals (CIs) are reported showing the level of change in outcome measures per unit increase in log-transformed LINE-1 DNA methylation, after adjustment for sex. Overall R^2 values for the models including both log-transformed LINE-1 DNA methylation and sex are given as percentages. Direct R^2 values for methylation were estimated by subtracting the R^2 value for the models including sex only from the corresponding models including both log-transformed LINE-1 DNA methylation and sex. This gives an estimate of the direct association between outcome and methylation after adjustment for any potential mediation through sex. Finally, the potential influence of covariates known to be associated with global DNA methylation (namely, age, alcohol consumption and smoking status) upon the observed associations was assessed within the linear regression models.

Ethical approval for the study was obtained from the appropriate local research ethics committees and all participants provided their written consent.

Results

LINE-1 DNA methylation measurements were available for 228 study members of the 412 who attended the clinical assessment, with no significant difference in the distribution of mean methylation values between males ($n = 85$, 37%) and females ($n = 143$, 63%) (Mann–Whitney $z = 0.60$, $P = 0.55$). Descriptive data for all variables used in this study are given in Table 1. Increased LINE-1 DNA methylation was associated with increasing fasting glucose [regression

coefficient (95% CI) = 2.80 (0.39–5.22) $P = 0.02$], total cholesterol = 4.76 (1.43–8.10), $P = 0.005$, total triglycerides = 3.83 (1.30–6.37), $P = 0.003$ and LDL cholesterol = 5.38 (2.12–8.64), $P = 0.001$ and with decreasing HDL cholesterol = −1.43 (−2.38 to −0.48), $P = 0.003$ and HDL:LDL ratio = −1.06 (−1.76 to −0.36), $P = 0.003$ (Table 2). For these analyses, LINE-1 DNA methylation (predictor variable) was log-transformed. Hence, these coefficients reflect the millimoles per litre change in outcome measure per unit increase in log-transformed methylation. Alternatively, following a 10% increase in LINE-1 DNA methylation, fasting glucose, total cholesterol, total triglycerides and LDL cholesterol increase by 0.28, 0.48, 0.38 and 0.54 mmol/l, respectively; and HDL cholesterol and HDL:LDL ratio decrease by 0.14 mmol/l and 0.11 units, respectively. The combined contributions of sex and LINE-1 DNA methylation explained between 4.98 and 9.65% of the variation in outcome measures. LINE-1 DNA methylation alone accounted for 2.12–4.37% of this variability (Table 2). Additional variables thought to influence global methylation (namely, age, alcohol consumption and smoking status) were not associated with LINE-1 DNA methylation in this study cohort and hence did not alter any of the associations observed (data not shown).

Table 1 Descriptive statistics for variables included in this investigation

Variable	*n*	Median (IQR)
Outcome		
BMI (kg/m^2)	228	25.70 (22.94–28.93)
Waist/hip ratio	228	0.84 (0.77–0.93)
Body fat (%)	226	40.35 (34.00–45.00)
Fasting glucose (mmol/l)	227	5.18 (4.90–5.50)
Total cholesterol (mmol/l)	228	5.14 (4.33–5.85)
Total triglycerides (mmol/l)	228	0.97 (0.70–1.52)
HDL cholesterol (mmol/l)	228	1.06 (0.85–1.30)
LDL cholesterol (mmol/l)	228	3.80 (2.91–4.58)
HDL:LDL ratio	228	0.28 (0.21–0.41)
Insulin secretion	213	14.7 (9.19–23.04)
HOMA-IR	219	1.68 (1.01–2.28)
Predictor		
Mean methylation (%)	228	52.76 (51.51–54.92)

IQR, interquartile range.

Discussion

The data presented demonstrate that increased LINE-1 DNA methylation is associated with a number of blood-based biomarkers of metabolic health and provide evidence of an association between LINE-1 DNA methylation and phenotypic traits other than cancer. There is substantial evidence that LINE-1 DNA methylation is modulated by a wide range of environmental exposures,[12–19] and several of these environmental exposures are associated with risk of complex diseases. It is therefore attractive to postulate that LINE-1 DNA methylation may provide a mechanistic link between such environmental exposures and the development of disease-related traits, although the current data suggest no such link with alcohol consumption and smoking status. However, it is also possible that LINE-1 DNA methylation patterns are confounded and are not causally

Table 2 Results of linear regression analyses of relationships between log-transformed methylation and the listed dependent variables, all adjusted for sex

Outcome variable	Coefficient (95% CI)	*P*-value	R^2 (%)	Direct R^2 (%)
BMI (kg/m^2)	2.30 (−9.99 to 14.59)	0.71	0.90	0.06
Waist/hip ratio	0.14 (−0.01 to 0.30)	0.07	65.42	0.51
Body fat (%)	5.14 (−19.01 to 29.30)	0.68	5.65	0.08
Fasting glucose (mmol/l)	2.80 (0.39 to 5.22)	0.02	9.05	2.12
Total cholesterol (mmol/l)	4.76 (1.43 to 8.10)	0.005	4.98	3.34
Total triglycerides (mmol/l)	3.83 (1.30 to 6.37)	0.003	9.65	3.57
HDL cholesterol (mmol/l)	−1.43 (−2.38 to −0.48)	0.003	8.96	3.54
LDL cholesterol (mmol/l)	5.38 (2.12 to 8.64)	0.001	7.02	4.37
HDL:LDL ratio	−1.06 (−1.76 to −0.36)	0.003	7.00	3.67
Insulin secretion	90.24 (−8.04 to 188.53)	0.07	1.57	1.54
HOMA-IR	2.34 (−2.68 to 7.36)	0.36	5.83	0.37

Coefficients and corresponding 95% CIs indicate the change in outcome measure per unit increase in log-transformed LINE-1 methylation, after adjustment for sex. R^2 reflects the variance (%) in outcome measures accounted for by both sex and log-transformed LINE-1 methylation (i.e. the combined effect of both covariates). Direct R^2 reflects the variance (%) in outcome measures accounted for by log-transformed LINE-1 methylation alone (i.e. the direct effect of methylation).

(or mechanistically) related to disease-related traits. Either way, there is considerable interest in the role of epigenetic mechanisms in common complex disease[29] given their potential to act as both informative diagnostic and prognostic biomarkers. It is postulated that, in those diseases with a prominent environmental component, it is possible that epigenetic factors contribute to the inter-individual differences in responses to environmental exposures[30] and to the pathogenesis of such diseases.[31]

We observed associations between LINE-1 DNA methylation and fasting concentrations of glucose, triglycerides and total, LDL and HDL cholesterol and also HDL:LDL ratio, all of which are blood-based biomarkers of increased risk of cardiovascular disease (CVD) and/or type 2 diabetes. High concentrations of fasting glucose are associated with the development of both CVD and diabetes.[32] We observed a positive association between LINE-1 DNA methylation and total cholesterol and triglycerides concentrations. Elevated concentrations of these blood lipid markers have been shown previously to be strongly associated with an increased risk of CVD.[33] Furthermore, our results show LINE-1 DNA methylation is associated with both increased LDL cholesterol and decreased HDL cholesterol concentrations. These opposing directional changes are those expected in individuals at increased risk of CVD. Given the high levels of collinearity of many of these measures, we chose however not to take a multivariable approach to the analysis.

As this study was conducted in individuals at age 50 years with no evidence of overt CVD, it was not possible to explore a potential association between LINE-1 DNA methylation and subsequent disease phenotype. However, this is a potential strength of this study as it removes the possible confounding effect of disease status on LINE-1 DNA methylation patterns. The NTFS birth cohort is being followed up longitudinally, which will provide the opportunity to ascertain the predictive utility of LINE-1 DNA methylation at age 50 years in respect of later disease risk. Nonetheless, given that both LINE-1 DNA methylation and blood biochemical measures were assessed at the same time-point, it is not possible to determine the direction of effect between these factors, if indeed, a direct causal (or mechanistic) link exists. Furthermore, given the small effect sizes observed, the contribution of one factor upon the other remains modest.

These findings are among the first observations to link LINE-1 DNA methylation levels with disease-related traits other than cancer. In the Boston-based Normative Aging Study, persons with prevalent IHD and stroke had, in contrast to our findings, lower LINE-1 DNA methylation and, in longitudinal analyses, those with lower LINE-1 DNA methylation were at higher risk for incident IHD, stroke and total mortality.[10] In a further study of the Boston-based cohort, an association was seen between LINE-1 hypomethylation and vascular cell adhesion molecule-1 for disease-free individuals, but not for those with prevalent IHD or stroke.[34] However, this all-male cohort is considerably older than the NTFS, with a mean age of 74 years at DNA sampling, which may offer some explanation as to the discordance in observations between the cohorts. The widely observed hypomethylation of LINE-1 DNA associated with cancer and the observations reported in the Normative Ageing Study could suggest that LINE-1 DNA methylation would be inversely associated with blood-based biomarkers of metabolic health, whereas our observations demonstrate the opposite association. Given the limited empirical data in this area and the lack of clear association between advancing age and decreased LINE-1 DNA methylation,[4,6] the current findings warrant further attention. There is some evidence that gene-specific DNA methylation is positively correlated with older age. Ronn *et al.*[35] showed that elderly, compared with young, non-diabetic twins had both higher DNA methylation and lower gene expression of *COX7A1* (a gene associated with peripheral insulin sensitivity, measured in 10 individuals using bisulphite sequencing). Hernandez *et al.*[36] also recently reported extensive evidence of genome-wide gene-specific hypermethylation with advancing age. The relationship between these gene-specific observations and global LINE-1 DNA methylation remains to be clarified.

Our observations suggest that LINE-1 DNA methylation and hence potentially other forms of epigenetic modification, might be useful in predicting risk of common complex diseases such as type 2 diabetes and CVD. The issues of confounding and reverse causation are fundamental to pursuing this further. DNA methylation is in essence a phenotype and is therefore vulnerable to multiple confounding influences including age, sex, smoking and socio-economic position to name only a few. Although our statistical appraisal of potential confounders did not highlight any obvious culprits, the issue cannot be dismissed. Indeed, it may transpire that DNA methylation provides nothing more than an indirect measure of confounding influences. In addition, it will be crucial to understand the causal relationship between LINE-1 DNA methylation and the blood-based biomarkers associated with this epigenetic signature, whether LINE-1 DNA methylation is causal in altering blood-based biomarkers such as fasting glucose and lipid concentrations or whether the reverse applies. There is limited evidence to suggest that altering glucose levels changes DNA methylation patterns,[37] but to our knowledge there is no direct evidence to link lipid levels to perturbed DNA methylation, or vice versa. A recent study of genome-wide methylation in cord-blood DNA highlighted numerous methylation-variable loci whose biological roles were related to lipid metabolism, suggesting a causal influence of altered methylation on

lipid levels.[19] A previous study of patients with coronary artery disease and controls showed that global DNA methylation was associated with coronary artery disease risk, and that this association was accentuated by increased plasma homocysteine concentration.[38] Further insight into the direction of causality may be obtained by adoption of a Mendelian randomization approach, as proposed recently by Relton and Davey Smith.[39] This approach involves the use of genetic variants as proxies for specific exposures, such that an association between genotype and DNA methylation would be indicative of a causal relationship (as lipid levels could not plausibly influence genotype and thus the possibility of reverse causation is removed). Numerous genetic variants have recently been reported to influence blood lipid profiles[40] and these could be used collectively as a proxy for lipid concentrations to investigate the association between lipid levels and DNA methylation.

In this study we estimated global DNA methylation using the LINE-1 assay, which measures cytosine methylation in common non-coding sequences that occur widely across the genome. The functional consequences of altered DNA methylation at these CpG sites within LINE-1 for the development of CVD and diabetes-related risk markers is not known and indeed may not be easily decipherable through the analysis of non-target tissues such as peripheral blood. A limitation of this, and many similar studies, is the reliance upon epigenetic profiling of peripheral blood DNA, with the assumption that it will be informative about target tissues.[41] Interrogation of methylation status of promoters in genes implicated directly in pathways of lipid metabolism and glucose homeostasis may provide greater insight. Methylation has been reported to change with both actual age (serial sampling)[42–44] and chronological age (cross sectional sampling).[36] As the study members were all born within a 2-month period in 1947 and assessed within an 18-month period at age 49–51 years, this minimizes the likelihood of confounding effects of chronological age.

In summary, we have presented evidence which supports the hypothesis that global LINE-1 DNA methylation at age 50 years is associated with biomarkers of metabolic health. Although these cross-sectional associations do not allow conclusions to be drawn with respect to the direction of causation, and the potential for confounding cannot be dismissed, the findings may have important implications for prediction, early diagnosis, prevention and treatment of common complex diseases such as CVD and type 2 diabetes.

Funding

MRC studentship (to J.C.Mc.); the Centre for Brain Ageing & Vitality (which is funded through the Lifelong Health and Wellbeing cross-council initiative by the MRC, BBSRC, EPSRC and ESRC); Biomedical Research Centre in Ageing and Age-Related Disease grant funded by the NHS Newcastle Hospitals Trust and the UK National Institute of Health Research.

Acknowledgements

The authors thank all Newcastle Thousand Family Study members for taking part in this study, study teams past and present and the various funding bodies that have contributed to funding this study since its inception. C.L.R. conceived and, with M.S.P., oversaw the project. L.P. oversaw the initial collection of samples and the clinical assessment. M.S.P. is the overall Director of the Newcastle Thousand Families Study. J.C.Mc. and L.M.B. analysed the samples, M.S.P. and C.P. analysed the data. M.S.P., J.C.M., J.C.Mc., C.P. and C.L.R. wrote the paper. M.S.P. and C.L.R. have primary responsibility for final content and act as guarantors for this article. All authors read and approved the final manuscript.

Conflict of interest: None declared.

KEY MESSAGES

- Patterns of both global and gene-specific DNA methylation change with age and these changes are believed to be associated with the development of common complex diseases.
- Associations were seen between global LINE-1 DNA methylation and a number of blood glucose and lipid markers (positive for fasting glucose, total cholesterol and triglycerides; negative for HDL cholesterol and the HDL:LDL ratio).
- Confounding and reverse causation represent major problems in epigenetic association studies and require careful consideration in studies of this type.
- These novel associations between global LINE-1 DNA methylation and blood glucose and lipid profiles highlight a potential role for epigenetic biomarkers as predictors of metabolic disease and may be relevant to future diagnosis, prevention and treatment of this group of disorders.

References

[1] Rakyan VK, Preis J, Morgan HD, Whitelaw E. The marks, mechanisms and memory of epigenetic states in mammals. *Biochem J* 2001;**356:**1–10.

[2] Waterland RA, Jirtle RL. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* 2003;**23:**5293–300.

[3] Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P. Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res* 2010;**38:**3909–22.

[4] Bollati V, Schwatz J, Wright R *et al*. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev* 2009;**130:**234–39.

[5] Wu HC, Delgado-Cruzata L, Flom JD *et al*. Global methylation profiles in DNA from different blood cell types. *Epigenetics* 2011;**6:**76–85.

[6] Ogino S, Nosho K, Kirkner GJ *et al*. A cohort study of tumoural LINE-1 hypomethylation and prognosis in colon cancer. *J Natl Cancer Inst* 2008;**100:**1734–38.

[7] Ahn JB, Chung WB, Maeda O *et al*. DNA methylation predicts recurrence from resected stage III proximal colon cancer. *Cancer* 2011;**117:**1847–54.

[8] Jintaridth P, Mutirangura A. Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. *Physiol Genomics* 2010; doi:10.1152/physiolgenomics.00146.2009 [Epub 9 February 2010].

[9] Schernhammer ES, Giovannucci E, Kawasaki T, Rosner B, Fuchs CS, Ogino S. Dietary folate, alcohol and B vitamins in relation to LINE-1 hypomethylation in colon cancer. *Gut* 2010;**59:**794–99.

[10] Baccarelli A, Wright R, Bollati V *et al*. Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiology* 2010;**21:**819–28.

[11] Wang L, Wang F, Guan J *et al*. Relation between hyomethylation of long interspersed nucleotide elements and risk of neural tube defects. *J Clin Nutr* 2010;**91:**1359–67.

[12] Mathers JC, Strathdee G, Relton CL. Induction of epigenetic alterations by dietary and other environmental factors. *Adv Genet* 2010;**71:**3–39.

[13] Wan YJ, Li YY, Xia W *et al*. Alterations in tumor biomarker GSTP gene methylation patterns induced by prenatal exposure to PFOS. *Toxicology* 2010;**274:**57–64.

[14] Breton CV, Byun HM, Wenten M, Pan F, Yang A, Gilliland FD. Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *Am J Respir Crit Care Med* 2009;**180:**462–67.

[15] Liu F, Killian JK, Yang M *et al*. Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. *Oncogene* 2010;**29:**3650–64.

[16] Wright RO, Schwartz J, Wright RJ *et al*. Biomarkers of lead exposure and DNA methylation within retrotransposons. *Environ Health Perspect* 2010;**118:**790–95.

[17] Pilsner JR, Hu H, Ettinger A *et al*. Influence of prenatal lead exposure on genomic methylation of cord blood DNA. *Environ Health Perspect* 2009;**117:**1466–71.

[18] Baccarelli A, Wright RO, Bollati V *et al*. Rapid DNA methylation changes after exposure to traffic particles. *Am J Resp Crit Care Med* 2009;**179:**572–78.

[19] Fryer AA, Emes RD, Ismail MK *et al*. Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. *Epigenetics* 2011;**6:**86–94.

[20] Pearce MS, Unwin NC, Parker L, Craft AW. Cohort profile: The Newcastle Thousand Families Study. *Int J Epidemiol* 2009;**38:**932–37.

[21] Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low density lipoprotein cholesterol in plasma without use of preparative ultracentrifuge. *Clin Chem* 1972;**18:**499–502.

[22] Pearce MS, Unwin NC, Relton CL, Alberti KGMM, Parker L. Lifecourse determinants of fasting and post-challenge glucose at age 49-51 years: The Newcastle Thousand Families Study. *Eur J Epidemiol* 2005;**20:**915–23.

[23] Pearce MS, Unwin NC, Parker L, Alberti KGMM. Life course determinants of insulin secretion and sensitivity at age 50 years: The Newcastle Thousand Families Study. *Diabetes Metab Res Rev* 2006;**22:**118–25.

[24] Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985;**28:**412–19.

[25] Phillips DIW, Clark PM, Hales CN, Osmond C. Understanding oral glucose tolerance: Comparison of glucose or insulin measurements during the oral glucose tolerance test with specific measurements of insulin resistance and insulin secretion. *Diabet Med* 1994;**11:**286–92.

[26] World Health Organization. *Organisation, Monitoring Trends and Determinants in Cardiovascular Disease Project. MONICA Manual, Part III*. Geneva, Switzerland: World Health Organization, 1990.

[27] Choi SH, Worswick S, Byun HM *et al*. Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer. *Int J Cancer* 2009;**125:**723–29.

[28] Bollati V, Baccarelli A, Hou L *et al*. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res* 2007;**67:**876–80.

[29] Feinberg AP. Epigenetics at the epicenter of modern medicine. *JAMA* 2008;**299:**1345–50.

[30] Mathers JC. Session 2: personalised nutrition. Epigenomics: a basis for understanding individual differences. *Proc Nutr Soc* 2008;**67:**390–94.

[31] Turan N, Katari S, Coutifaris C, Sapienza C. Explaining inter-individual variability in phenotype: is epigenetics up to the challenge? *Epigenetics* 2010;**5:**16–19.

[32] Nathan DM, Davidson MB, DeFronzo RA *et al*. Impaired fasting glucose and impaired glucose tolerance: implications for care. *Diab Care* 2007;**30:**753–59.

[33] Sarwar N, Danesh J, Eiriksdottir G *et al*. Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. *Circulation* 2007;**115:**450–58.

[34] Baccarelli A, Tarantini L, Wright RO *et al*. Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study. *Epigenetics* **5;** doi:10.4161/epi.5.3.11377 [Epub 1 April 2010].

[35] Rönn T, Poulsen P, Hansson O *et al*. Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle. *Diabetologia* 2008;**51:**1159–68.

[36] Hernandez DG, Nalls MA, Gibbs JR *et al*. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet* 2011;**20:**1164–72.

[37] Sharma P, Kumar J, Garg G *et al*. Detection of altered global DNA methylation in coronary artery disease patients. *DNA Cell Biol* 2008;**27:**357–65.
[38] Li Y, Liu L, Tollefsbol TO. Glucose restriction can extend normal cell lifespan and impair precancerous cell growth through epigenetic control of hTERT and p16 expression. *FASEB J* 2010;**24:**1442–53.
[39] Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention and treatment. *PLoS Med* 2010;**7:**e1000356.
[40] Inouye M, Silander K, Hamalainen E *et al*. An immune response network associated with blood lipid levels. *PLoS Genet* 2010;**6:**e1001113.
[41] McKay JA, Xie L, Harris S, Wong YK, Ford D, Mathers JC. Blood as a surrogate marker for tissue-specific DNA methylation and changes due to folate depletion in post-partum female mice. *Mol Nutr Food Res* 2011;**55:**1026–35.
[42] Bjornsson HT, Sigurdsson MI, Fallin MD *et al*. Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 2008;**299:**2877–93.
[43] Maegawa S, Hinkal G, Kim HS *et al*. Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res* 2010;**20:**332–40.
[44] Rakyan VK, Down TA, Maslau S *et al*. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 2010;**20:**434–39.

Gene 499 (2012) 99–107

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Bioinformatic selection of putative epigenetically regulated loci associated with obesity using gene expression data

Valérie Turcot [a,1], Alexandra Groom [a], James C. McConnell [a], Mark S. Pearce [b], Catherine Potter [a], Nicholas D. Embleton [c], Daniel C. Swan [d], Caroline L. Relton [a,*]

[a] Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK
[b] Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK
[c] Neonatal Service, Royal Victoria Infirmary, Newcastle upon Tyne, UK
[d] Bioinformatics Unit, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, UK

ARTICLE INFO

Article history:
Accepted 1 February 2012
Available online 9 March 2012

Keywords:
DNA methylation
CpG island
Promoter *in silico* analysis
Body mass index
Adiposity

ABSTRACT

There is considerable interest in defining the relationship between epigenetic variation and the risk of common complex diseases. Strategies which assist in the prioritisation of target loci that have the potential to be epigenetically regulated might provide a useful approach in identifying concrete examples of epigenotype–phenotype associations. Focusing on the postulated role of epigenetic factors in the aetiopathogenesis of obesity this report outlines an approach utilising gene expression data and a suite of bioinformatic tools to prioritise a list of target candidate genes for more detailed experimental scrutiny. Gene expression microarrays were performed using peripheral blood RNA from children aged 11–13 years selected from the Newcastle Preterm Birth Growth Study which were grouped by body mass index (BMI). Genes showing ≥2.0 fold differential expression between low and high BMI groups were selected for *in silico* analysis. Several bioinformatic tools were used for each following step; 1) a literature search was carried out to identify whether the differentially expressed genes were associated with adiposity phenotypes. Of those obesity-candidate genes, putative epigenetically regulated promoters were identified by 2) defining the promoter regions, 3) then by selecting promoters with a CpG island (CGI), 4) and then by identifying any transcription factor binding modules covering CpG sites within the CGI. This bioinformatic processing culminated in the identification of a short list of target obesity-candidate genes putatively regulated by DNA methylation which can be taken forward for experimental analysis. The proposed workflow provides a flexible, versatile and low cost methodology for target gene prioritisation that is applicable to multiple species and disease contexts.

1. Introduction

Overweight and obesity are graded conditions of excess body fat, which are clinically defined based on their associated risk for comorbidities and mortality in the adult population (NIH Report, 1998; WHO, 2006). The total direct cost of overweight and obesity to the National Health Service (NHS) in the UK was estimated at £5.15 billion in 2006–2007, corresponding to 16.2% of NHS total costs in the same period which was primarily due to hypertensive disease, osteoarthritis, diabetes mellitus and ischemic heart disease (Scarborough et al., 2011). In children and adolescents, the evaluation of age and sex specific body mass index (BMI) has been shown to be the most useful method for assessing weight status and identifying those at a higher risk of future adverse health outcomes (Cole et al., 2000; Janssen et al., 2005; Kuczmarski et al., 2000; WHO, 2006). The prevalence of overweight and obesity in children and adolescents is increasing in England (Stamatakis et al., 2010a, 2010b), which renders them to be at a higher risk of becoming overweight adults (Singh et al., 2008). This situation may increase the burden of overweight and obesity in future years and thus prompts the need for prevention and therapeutic interventions. To achieve this goal a better knowledge of the contributing factors is essential.

A number of epidemiological studies and animal models have shown that maternal health and nutritional status during gestation and lactation have long-term effects on systems regulating energy

Abbreviations: aa, amino acid; BMI, body mass index; bp, base pair; CD38, cluster of differentiation 38; CDC42, cell division cycle 42; CGI, CpG island; CpG, cytosine–phosphate–guanine; FDR, false discovery rate; IOTF, International Obesity Task Force; LTF, lactotransferrin; NCBI, National Center for Biotechnology Information; TACSTD2, tumour-associated calcium signal transducer 2; TFBM, transcription factor binding module; TFBS, transcription factor binding site; TLSS, translation start site; UTR, untranslated region.

* Corresponding author at: Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK. Tel.: +44 191 2418623; fax: +44 191 2418666.
E-mail address: caroline.relton@ncl.ac.uk (C.L. Relton).
[1] Permanent address: Institute of Nutraceuticals and Functional Foods, Pavillon des Services, Université Laval, 2440 Hochelaga Blvd., Quebec City, QC, Canada, G1V 0A6.

0378-1119/$ – see front matter © 2012 Published by Elsevier B.V.
doi:10.1016/j.gene.2012.02.001

balance in the developing offspring (reviewed by Sullivan and Grove, 2010). The molecular mediators of early metabolic programming of obesity in offspring are poorly understood, but may partly implicate long term disruption of glucose, insulin, leptin and inflammatory cytokine homeostasis, as well as epigenetic mechanisms (Sullivan and Grove, 2010). Potential interactions between the environment and epigenetics, particularly in periods of high developmental plasticity in early life, may mediate the expression of genes associated with increased BMI and adiposity which could partly explain the inter-individual differences in obesity risk (Campion et al., 2009; Herrera et al., 2011). The logical causal pathway would be that exposure (*i.e.* environmental factors) influences epigenetic patterns which in turn changes the expression of genes implicated in the etiology of obesity (Relton and Davey, 2010). However, any association linking epigenetic variation to obesity may be vulnerable to confounding and reverse causation (where the obese state might alter the epigenome and not vice versa) (Relton and Davey, 2010; Schadt et al., 2005). This situation supports the necessity to pursue investigations to better understand the relationship between exposure, epigenetic patterns and complex diseases, such as obesity, in order to evaluate the utility of treating the disease via epigenetic-based interventions or using epigenetic patterns as a diagnostic tool (Relton and Davey, 2010).

DNA methylation is the most widely studied epigenetic modification in humans which occurs mainly through the addition of a methyl group (CH_3) to a cytosine positioned next to a guanine nucleotide (CpG site). CpG sites tend to cluster together in regions called CpG islands (CGIs). Approximately 60% of human gene promoters are associated with CGIs that are usually unmethylated in normal cells (*i.e.* non-tumourigenic cells). Methylated DNA can inhibit gene expression by various mechanisms, such as promoting the recruitment of methyl-CpG-binding domain (MBD) proteins which has a downstream effect on the ability of transcription factors to access their target sites (Portela and Esteller, 2010). Interestingly, some genes previously associated with obesity (Rankinen et al., 2006) have been shown to be epigenetically regulated, such as peroxisome proliferator-activated receptor gamma (*PPARG*), glucocorticoid receptor (*NR3C1*), leptin (*LEP*), lipoprotein lipase (*LPL*) and caveolin-1 (*CAV1*) (reviewed by Campion et al., 2009). Some of these candidate genes have a promoter CGI suggesting that their expression may be regulated by DNA methylation and could potentially explain inter-individual differences in obesity risk (Campion et al., 2009). To date, there is limited evidence linking epigenetic variability in specific genes with common complex disease phenotypes including obesity. This situation may be explained by the fact that adequately powered studies relating epigenetic profiles and disease-related traits are few in number (Relton and Davey, 2010). Although recently, tangible evidence supports the association between DNA methylation pattern and adiposity phenotypes in humans (Godfrey et al., 2011; Stepanow et al., 2011), which underlines the necessity to conduct studies that will help to discover relevant genes that may be epigenetically regulated by DNA methylation and are associated with overweight and obesity.

A number of potential methodological approaches exist whereby loci can be identified for prioritisation for epigenetic investigation, including a variety of data mining and bioinformatics approaches. Where epigenetic variation contributes to inter-individual variation in gene expression and thus to variation in common complex disease risk, gene expression microarrays provide a useful tool to identify genes differentially expressed between variable adiposity phenotypes. But further strategies are needed to specifically target obesity-candidate genes that may be epigenetically controlled by DNA methylation. Bioinformatic approaches for the prioritisation of epigenetic target genes have previously been described in the context of differential gene expression response following nutritional exposure *in utero* (McKay et al., 2008). Inspired by this approach, this paper proposes a refined multistep *in silico* analysis using bioinformatic tools to identify a list of prioritised genes for further experimental analyses that have the potential to be specifically associated with obesity development and for which their expression may be regulated by DNA methylation in their promoter CGI using data from gene expression microarrays in children grouped by BMI.

2. Materials and methods

2.1. Study populations

The gene expression analysis was conducted on children aged 11–13 years selected from the Preterm Birth Growth Study, recruited by the Special Care Baby Unit, Royal Victoria Infirmary, Newcastle upon Tyne, UK (Cooke et al., 1999; Cooke et al., 2001), who participated in a follow-up clinical examination of cardiometabolic traits during 2007–2008. They were all healthy preterm infants with no evidence of systemic disease, required no medication, and were growing normally at the time of hospital discharge. Anthropometric and body composition data using a dual energy X-ray absorptiometry were taken at 11–13 years of age. Gene expression data were available for 24 children in this cohort who were divided into tertiles according to BMI. The children in the upper and lower tertiles ($n=7$ per group) were compared for the purposes of this experiment. Summary details of the two groups are provided in Table 1. There was no significant difference in height or age between the two groups. As expected, they were significantly discordant in body weight, BMI and fat mass ($P \leq 0.0006$). Based on the proposed age and sex specific BMI cut off points for overweight and obese children from the International Obesity Task Force (IOTF) (Cole et al., 2000), all the children in the low BMI group had a "healthy" BMI. Alternatively, those in the high BMI group were considered as overweight ($n=5$) or obese ($n=1$) except for one children with a limit "healthy" BMI. This study was approved by the Ethics Committee of the Newcastle and North Tyneside Health Authority, and informed consent was obtained from the parent(s) or legal guardian.

2.2. Preparation of nucleic acid

A volume of 2.5 ml of peripheral blood was drawn into a PAXgene™ Blood RNA tube (PreAnalytiX QIAGEN GmbH, Affymetrix Inc., Santa Clara, California), incubated at room temperature for 2 h and then stored at −70 °C until extraction. Total RNA was extracted using the PAXgene™ Blood RNA System Kit following the manufacturer's instructions. RNA Integrity Number was assessed using RNA Nano 6000 chips run on an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Palo Alto, California, USA) and concentration determined using a NanoDrop™ ND-1000 spectrophotometer (NanoDrop Technologies, Thermo Fisher Scientific Inc., Waltham, MA, USA).

2.3. Gene expression analysis

RNA samples were sent to ServiceXS (Leiden, The Netherlands) for globin reduction, labelling, hybridization to Human NuGO-

Table 1
Comparison of low and high BMI groups selected from the Preterm Birth Growth Study for gene expression analysis.

Variable	Low BMI (n = 7)	High BMI (n = 7)	*P*-value
Female (n)	3	3	–
Height (cm)	150.1 (10.0)	150.7 (11.7)	0.911
Weight (kg)	35.4 (5.9)	59.1 (12.1)	0.0006
Body mass index (kg/m^2)	15.6 (0.7)	25.8 (2.6)	<0.0001
Age (months)	144.7 (10.8)	151.7 (11.8)	0.268
Age term adjusted (months)	134.5 (13.1)	142.3 (12.4)	0.276
Fat mass (kg)	9.5 (3.7)	24.0 (4.7)	<0.0001
Lean mass (kg)	25.0 (3.7)	32.7 (6.8)	0.022

Mean (standard deviation) values are presented in this table with t-test or Mann–Whitney U test statistics for between group comparisons.

Hs1a520180 GeneChip arrays and scanning of the arrays. Globin reduction was performed using GeneChip® Globin-Reduction kit (PreAnalytiX QIAGEN GmbH, Affymetrix Inc., Santa Clara, California) according to the manufacturer's instructions (Mat. no. 1029528) using Peptide Nucleic Acid oligonucleotides complementary to human globin mRNA transcripts (GR PNA-L G2001 Panagene Inc., Korea) and Globin-Reduction RNA controls (No. 900586, PreAnalytiX QIAGEN GmbH, Affymetrix Inc., Santa Clara, California). Since total RNA was purified from whole blood, it contains high amounts of globin transcripts (in contrast to fractionated blood samples). Globin reduction was thus necessary to reduce the amount of cDNA generated from globin mRNA during reverse transcription, enabling sensitive and unbiased gene expression analysis. Human NuGO-Hs1a520180 GeneChip CEL files were normalized in BioConductor (Gentleman et al., 2004) (http://www.bioconductor.org) using the GeneChip Robust Multi-array Average (GCRMA) procedure as implemented in the gcrma package. Genes with differential expression between BMI groups were identified with the RankProd package (False Discovery Rate (FDR) value <0.05 with 100 permutations of the class labels) (Hong et al., 2006) (http://www.bioconductor.org.). Annotations were attached to probe sets from the nugohs1a520180.db library (http://www.bioconductor.org/help/bioc-views/2.6/data/annotation/html/nugohs1a520180.db.html). Raw and normalized data from the experiment was deposited in GEO (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE22013. For the *in silico* analysis genes which were differentially expressed ≥2.0 fold in low *vs*. high BMI and high *vs*. low BMI were prioritised for further analyses. We used this cut off point to firstly prioritise the genes that were more greatly differentially expressed, which would result in a fewer number of obesity-candidate genes putatively regulated by DNA methylation as a first exploratory investigation.

2.4. Literature search

To identify genes differentially expressed ≥2.0 fold that may be physiologically relevant in the development of obesity we utilised the Genomatix tool LitInspector (Frisch et al., 2009) (Genomatix Inc., Munich, German; http://www.genomatix.de). This program is a literature search tool providing gene and signal transduction pathway mining within the National Center for Biotechnology Information (NCBI)'s PubMed database. The main advantage of this program compared to a PubMed search is that it speeds up the literature research by increasing the readability of abstracts using an automatic gene recognition and color coding of associated diseases and submitted keywords, and also provides a shorter list of relevant abstracts due to its high gene recognition performance. Its gene recognition is based on the comprehensive gene synonym lists provided by the NCBI's Entrez Gene, thus the submission of a single synonym will consider all synonyms of this gene, and the program has a high recognition quality explained by its ability to resolve homonyms and ambiguous synonyms, as well as rejecting "non-gene" abbreviations. This program has been fully described previously (Frisch et al., 2009). In this report the gene symbol for each targeted differentially expressed locus was used as the gene name identifier and each keyword related to adiposity phenotypes (body mass index, obesity, fat mass, adipose tissue) was added as free text. The keyword methylation was also used to verify whether DNA methylation variability had previously been observed in these genes. Genes, or their encoded protein, which had previously been associated with an adiposity phenotype in the literature were retained for promoter region analysis. As demonstrated by several users, the LitInspector software had a greater gene recognition performance compared to other text mining tools which can identify co-occurring gene names and custom keywords within the PubMed database (*e.g.* PolySearch (Cheng et al., 2008) and iHOP (Hoffmann and Valencia, 2005)) (Frisch et al., 2009). PubMatrix (Becker et al., 2003) would have provided an alternative tool but, to the best of our knowledge, its gene recognition performance has not been previously evaluated.

2.5. Promoter region analysis

From the list of differentially expressed genes potentially associated with obesity we selected only those with a promoter region with the potential to be regulated by DNA methylation. To do so, we firstly identified the promoter region of the candidate genes using the Genomatix tool Gene2Promoter which gives the promoter sequences of all alternative transcripts for a given locus. Putative promoter sequences of validated transcripts with a Reference Sequence (RefSeq) recorded in the NCBI (http://www.ncbi.nlm.nih.gov/RefSeq/) public database were downloaded in FASTA format for each obesity candidate gene. These promoter sequences were then submitted to CpG Island Explorer software (http://bioinfo.hku.hk/cpgieintro.html) to determine which transcripts have a putative promoter containing a CGI. The NCBI Map Viewer database was also used (Human Build 37.2; http://www.ncbi.nlm.nih.gov/projects/mapview/) to delimit CGIs that may extend beyond the putative promoter sequences submitted to CpG Island Explorer. The Takai and Jones algorithm (Takai and Jones, 2002) was used for the detection of a promoter CGI which corresponded to 200 bp minimum length, a G + C content ≥50%, a ratio of observed CpG/expected CpG sites ≥0.60, and where islands ≤100 base pairs (bp) apart were merged. Since methylation

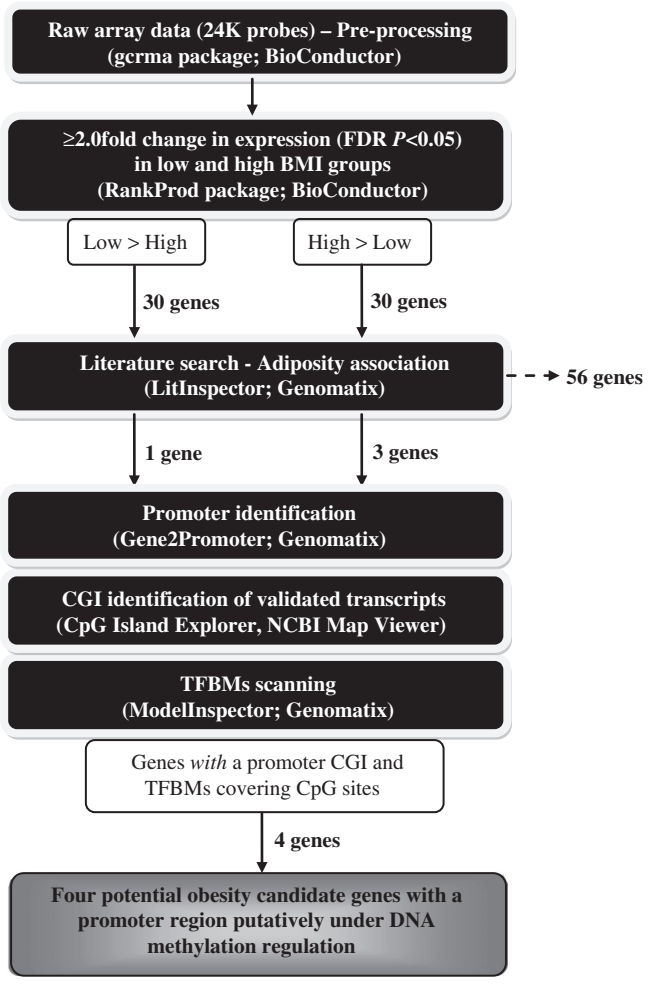


Fig. 1. Overview of the multistep *in silico* analysis performed for the selection of the obesity candidate genes putatively regulated by DNA methylation. Unsuitable and suitable genes for the next step of analysis are depicted with dotted and plain lines respectively. BMI: body mass index, CGI: CpG island, FDR: false-discovery rate, TFBM: transcription factor binding module, NCBI: National Center for Biotechnology Information.

of CpG sites within the promoter may affect the ability of transcription factors to access their target site and influence the gene expression regulation (Portela and Esteller, 2010) we submitted the sequences of the putative promoters containing CGI to the Genomatix tool ModelInspector (Klingenhoff et al., 1999). This program searches for transcription factors binding modules (TFBM) that contain at least two transcription factor binding sites (TFBS) in a functionally defined distance range in the submitted sequences (for more details about TFBM definition, refer to Klingenhoff et al., 1999). The Vertebrate Module Library version 5.3 was selected for the analysis and only the TFBMs identified in the *Homo sapiens* organism were considered in the final results. The putative promoters containing CGIs with at

Table 2
Genes ≥2.0 fold differentially expressed in peripheral blood of children aged 11–13 years from the Preterm Birth Growth Study grouped in low and high BMI.

Gene symbol	Gene name	Affymetrix probe set	Fold change	FDR *P*-value
Genes ≥2.0 fold more expressed in low BMI children; n = 30 genes				
APOBEC3B	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B	206632_s_at	3.71	$<1\times10^{-5}$
RRM2	Ribonucleotide reductase M2	201890_at	3.15	2×10^{-3}
N.A.	Hs.19156	1555989_at	2.75	1.1×10^{-3}
TYMS	Thymidylate synthetase	202589_at	2.72	2×10^{-3}
CD38	Cluster of differentiation 38	205692_s_at	2.69	$<1\times10^{-5}$
IGHA1	Immunoglobulin heavy constant alpha 1	217022_s_at	2.67	$<1\times10^{-5}$
IGKV6-21	Immunoglobulin Kappa light chain V gene segment	NuGO_eht0328018_s_at	2.62	1×10^{-3}
KIR3DL2	Killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 2	207314_x_at	2.35	8×10^{-4}
SERPING1	Serpin peptidase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	200986_at	2.33	2×10^{-3}
IFI44L	Interferon-induced protein 44-like	204439_at	2.30	$<1\times10^{-5}$
KIR3DL2	Killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 2	207313_x_at	2.28	$<1\times10^{-5}$
CCNA2	Cyclin A2	213226_at	2.28	$<1\times10^{-5}$
MGC29506	Hypothetical protein MGC29506	223565_at	2.27	$<1\times10^{-5}$
S100B	S100 calcium binding protein B	209686_at	2.27	$<1\times10^{-5}$
CCNB2	Cyclin B2	202705_at	2.24	$<1\times10^{-5}$
FAM72D	Family with sequence similarity 72, member D	225834_at	2.23	$<1\times10^{-5}$
IGLV1-51	Immunoglobulin lambda variable 1-51	217179_x_at	2.18	4×10^{-4}
UHRF1	Ubiquitin-like, containing PHD and RING finger domains, 1	225655_at	2.15	6×10^{-3}
TRBV4-2	T-cell receptor beta V gene segment	NuGO_eht0332013_s_at	2.14	$<1\times10^{-5}$
OAS1	2′,5′-oligoadenylate synthetase 1, 40/46 kDa	202869_at	2.13	4×10^{-4}
IGKC	Immunoglobulin kappa constant	NuGO_eht0355658_x_at	2.11	$<1\times10^{-5}$
DTL	Denticleless homolog (*Drosophila*)	218585_s_at	2.08	3×10^{-3}
IGLV1-44	Immunoglobulin lambda variable 1-44	234764_x_at	2.06	$<1\times10^{-5}$
PLGLB2	Plasminogen-like B2	205871_at	2.05	6×10^{-3}
DUSP5	Dual specificity phosphatase 5	209457_at	2.05	$<1\times10^{-5}$
CPA3	Carboxypeptidase A3 (mast cell)	205624_at	2.04	$<1\times10^{-5}$
IGKV1D-16	Immunoglobulin kappa light chain V gene segment	NuGO_eht0241620_x_at	2.03	$<1\times10^{-5}$
BUB1	BUB1 budding uninhibited by benzimidazoles 1 homolog (yeast)	209642_at	2.02	$<1\times10^{-5}$
ERAP2	Endoplasmic reticulum aminopeptidase 2	227462_at	2.01	$<1\times10^{-5}$
BCAT1	Branched chain aminotransferase 1, cytosolic	225285_at	2.01	5×10^{-4}
Genes ≥2.0 fold more expressed in high BMI children; n = 30 genes				
HLA-DQA1	Major histocompatibility complex, class II, DQ alpha 1	213831_at	5.61	$<1\times10^{-5}$
KRT1	Keratin 1 (epidermolytic hyperkeratosis)	205900_at	4.35	$<1\times10^{-5}$
CDC42	Cell division cycle 42 (GTP binding protein, 25 kDa)	208727_s_at	3.86	$<1\times10^{-5}$
SLPI	Secretory leukocyte peptidase inhibitor	203021_at	3.50	$<1\ x10^{-5}$
HLA-DRB1,3,4	HLA-DRB1 HGNC major histocompatibility complex, class II, DR beta 1	238900_at	3.40	2×10^{-3}
MYOM2	Myomesin (M-protein) 2165 kDa	205826_at	2.84	$<1\times10^{-5}$
TUBB2A	Tubulin, beta 2A	204141_at	2.80	1×10^{-3}
CHST13	Carbohydrate (chondroitin 4) sulfotransferase 13	239647_at	2.67	$<1\times10^{-5}$
LTF	Lactotransferrin	202018_s_at	2.62	$<1\times10^{-5}$
S100P	S100 calcium binding protein P	204351_at	2.58	$<1\times10^{-5}$
COL9A3	Collagen, type IX, alpha 3	NuGO_eht0343916_s_at	2.57	1×10^{-3}
TMOD1	Tropomodulin 1	203661_s_at	2.53	1×10^{-3}
ANXA3	Annexin A3	209369_at	2.50	$<1\times10^{-5}$
RFX2	Regulatory factor X, 2 (influences HLA class II expression)	226872_at	2.46	$<1\times10^{-5}$
S100A12	S100 calcium binding protein A12	205863_at	2.44	$<1\times10^{-5}$
TACSTD2	Tumour-associated calcium signal transducer 2 precursor	202286_s_at	2.40	2×10^{-3}
SLC4A1	Solute carrier family 4, anion exchanger, member 1 (erythrocyte membrane protein band 3, Diego blood group)	205592_at	2.36	1×10^{-3}
SNCA	Synuclein, alpha (non A4 component of amyloid precursor)	236081_at	2.29	2×10^{-3}
FECH	Ferrochelatase (protoporphyria)	203115_at	2.26	1×10^{-3}
SELENBP1	Selenium binding protein 1	214433_s_at	2.25	2×10^{-3}
AMFR	Autocrine motility factor receptor	202203_s_at	2.21	1×10^{-3}
TMTC1	Transmembrane and tetratricopeptide repeat containing 1	226931_at	2.19	2×10^{-3}
PGLYRP1	Peptidoglycan recognition protein 1	NuGO_eht0008938_at	2.18	2×10^{-3}
NFXL1	Nuclear transcription factor, X-box binding-like 1	227220_at	2.15	1×10^{-3}
ALPL	Alkaline phosphatase, liver/bone/kidney	215783_s_at	2.12	1×10^{-3}
PI3	Probable ATP-dependent DNA helicase HFM1 (EC 3.6.1.) (SEC63 domain- containing protein 1)	41469_at	2.11	9×10^{-4}
DEFA4	Defensin, alpha 4, corticostatin	207269_at	2.09	2×10^{-3}
DEFA1	Defensin, alpha 1	205033_s_at	2.04	1×10^{-3}
N.A.	Hs.606581	233217_at	2.02	9×10^{-4}
CMTM2	CKLF-like MARVEL transmembrane domain containing 2	229967_at	2.01	1×10^{-3}

BMI: body mass index; FDR: false discovery rate; N.A.: not available.

least one TFBM covering one or multiple CpG sites were considered as putatively regulated by DNA methylation.

Globally, this multi-step *in silico* analysis (Sections 2.4 and 2.5) would select a proportion of differentially expressed obesity candidate genes likely to be under epigenetic regulation for further DNA methylation quantification and adiposity-association analyses. As also discussed later in this report, the proposed approach will not identify *all* putatively epigenetically regulated genes but merely provides a valuable and efficient mode of prioritisation which utilises several *in silico* resources.

2.6. Statistical analysis

Clinical characteristics between children in the low and high BMI groups were compared using a Student's t-test for normally distributed variables or Mann–Whitney U for variables with a skewed distribution.

3. Results

The overview of the gene expression and *in silico* analysis workflow is depicted in Fig. 1. Each step describes which program was used and how many candidate genes were selected for downstream analyses. The results obtained for each step are described below.

3.1. Gene expression analysis

After data normalization and pre-processing 60 transcripts (corresponding to 60 genes) were differentially expressed ≥2.0 fold between low and high BMI groups (Table 2). Children with lower BMI had increased expression of 30 genes in comparison to children with higher BMI. Conversely, in children with higher BMI, the number of genes overexpressed compared to low BMI children was also 30.

3.2. Literature search

From the 60 differentially expressed genes (≥2.0 fold), a literature search was performed to identify those that were potentially obesity-related genes. Table 3 shows the genes for which an association with at least one adiposity phenotype (body mass index, obesity, fat mass and adipose tissue) has been documented in the literature as well as any previous associations with DNA methylation.

Among the transcripts that showed greater expression in RNA extracted from the peripheral blood of low BMI children, only one gene, cluster of differentiation 38 (*CD38*; 2.69 fold), had previously been associated with adiposity phenotypes. This gene codes for a transmembrane enzyme implicated in signal transduction and calcium signalling (Chini, 2009). It is a key enzyme in the control of intra- and extracellular nicotinamide–adenine dinucleotide levels (Aksoy et al., 2006a, 2006b; Chini, 2009). This protein was shown to be associated with obesity development in *cd38*-deficient mice on a high-fat diet and its role may be related to the energy expenditure regulation (Barbosa et al., 2007).

Among the transcripts with greater expression in peripheral blood of high BMI children, three corresponding genes were previously associated with adiposity phenotypes; these are cell division cycle 42 (*CDC42*; 3.86 fold): lactotransferrin (*LTF*; 2.62 fold): tumour-associated calcium signal transducer 2 (*TACSTD2*; 2.40 fold). The *CDC42* gene encodes a guanosine triphosphate (GTP)-binding protein member of the Rho GTPases family. It is mainly implicated in cytoskeleton organisation, polarity, migration, cell division and morphogenesis. CDC42 can mediate insulin signalling in the 3T3-L1 adipocyte cell line (Usui et al., 2003). Higher *Cdc42* expression levels in visceral adipose tissue were seen in animals on a high-fat diet, which may imply a possible role for CDC42 in fat accumulation (Hishikawa et al., 2005). Obesity is frequently associated with increased leptin levels and Jaffe et al. demonstrated a direct and dose- and time-dependant activation of the *CDC42* gene by leptin in aggressive human colon cancer cell lines (Jaffe and Schwartz, 2008).

LTF encodes a non-heme iron-binding protein and is part of the transferrin protein family. It is a major component of iron homeostasis regulation and the mammalian innate immune system. *Ltf* was shown to be expressed 1.6 fold greater in fat depots of obese versus lean BSB mice (model for complex obesity; backcross mice: (C57BL/6J × *Mus spretus*) × C57BL/6J) (Farahani et al., 2004). *In vitro* experiments revealed that the LTF protein may be implicated in adipogenesis, cell differentiation and adipose tissue integrity (Moreno-Navarrete et al., 2009). Yagi et al. noted a reduction of adipogenic differentiation and lipid droplets in a lactotransferrin-treated mouse preadipocyte cell line with a concomitant increase in cell number (Yagi et al., 2008). Addition of lactotransferrin improved cell viability in the media of visceral adipose explants from severely obese subjects (Fernandez-Real et al., 2010). Several studies also reported that *LTF* expression correlated with the methylation levels of CpG sites localized within and surrounding its promoter region (Grant et al., 1999; Shaheduzzaman et al., 2007; Teng et al., 2004).

The *TACSTD2* gene encodes a cell surface glycoprotein for which the main function remains largely unknown (Ibragimova et al., 2010). The cross linking of TACSTD2 with antibodies causes a transient increase in intracellular calcium levels and it may thus have a role in signal transduction (Fornaro et al., 1995). Previous experiments in our lab have shown a novel association between DNA methylation levels within the *TACSTD2* promoter CGI and fat mass content in children (Groom et al., 2012). Another study has also shown an

Table 3
List and main functions of the differentially expressed genes associated with adiposity phenotypes using the bioinformatic tool LitInspector.

Gene symbol	Gene name	Main function[a]	DNA methylation and adiposity phenotype associations
Gene ≥2.0 fold more expressed in low BMI children			
CD38	Cluster of differentiation 38	Signal transduction Calcium signaling	Obesity (Barbosa et al., 2007) DNA methylation (Ferrero et al., 1999)
Genes ≥2.0 fold more expressed in high BMI children			
CDC42	Cell division cycle 42 (GTP binding protein, 25 kDa)	Signal transduction GTPase activity	Adipose tissue (Hishikawa et al., 2005; Usui et al., 2003) Obesity (via leptin) (Jaffe and Schwartz, 2008)
LTF	Lactotransferrin	Iron ion homeostasis immune system	Obesity (Farahani et al., 2004) Adipose tissue (Fernandez-Real et al., 2010; Moreno-Navarrete et al., 2009; Yagi et al., 2008) DNA methylation (Grant et al., 1999; Shaheduzzaman et al., 2007; Teng et al., 2004)
TACSTD2	Tumour-associated calcium signal transducer 2	Signal transduction Calcium signaling	Fat mass (Groom et al., 2012) DNA methylation (Groom et al., 2012; Ibragimova et al., 2010; Jeronimo and Esteller, 2010)

BMI: body mass index.
[a] Relative to EntrezGene (http://www.ncbi.nlm.nih.gov/gene) and Gene Ontology (http://amigo.geneontology.org) databases.

association between the *TACSTD2* promoter CGI methylation levels and its expression in primary prostate tumour tissues in humans (Ibragimova et al., 2010; Jeronimo and Esteller, 2010).

3.3. Promoter region analysis

Analysis of the promoter region was conducted for each of the 4 obesity-candidate genes using a number of bioinformatic tools (Fig. 1) to identify those putatively regulated by DNA methylation. An overview of the results is described below and depicted in Fig. 2. The *CD38* gene located at the 4p15 locus codes for one validated transcript (RefSeq ID: NM_001775.2) that spans 1494 bp and contains 8 exons. It encodes a functional protein of 300 amino acids (aa). The Gene2Promoter software identified a putative promoter of 759 bp (Fig. 2A) covering 640 bp before and 119 bp after the translation start site (TLSS; relative to the A of the ATG-translation initiation codon). There was a CGI covering the last 199 bp of the putative promoter and 727 bp beyond this promoter. Submission of the promoter sequence to ModelInspector identified 8 TFBMs 4 of which had CpG sites within their binding sequence (see more details in Supplementary Table S1). The *CDC42* gene is located at the 1p36.1 locus and codes for three validated transcripts (RefSeq IDs: NM_001039802.1,

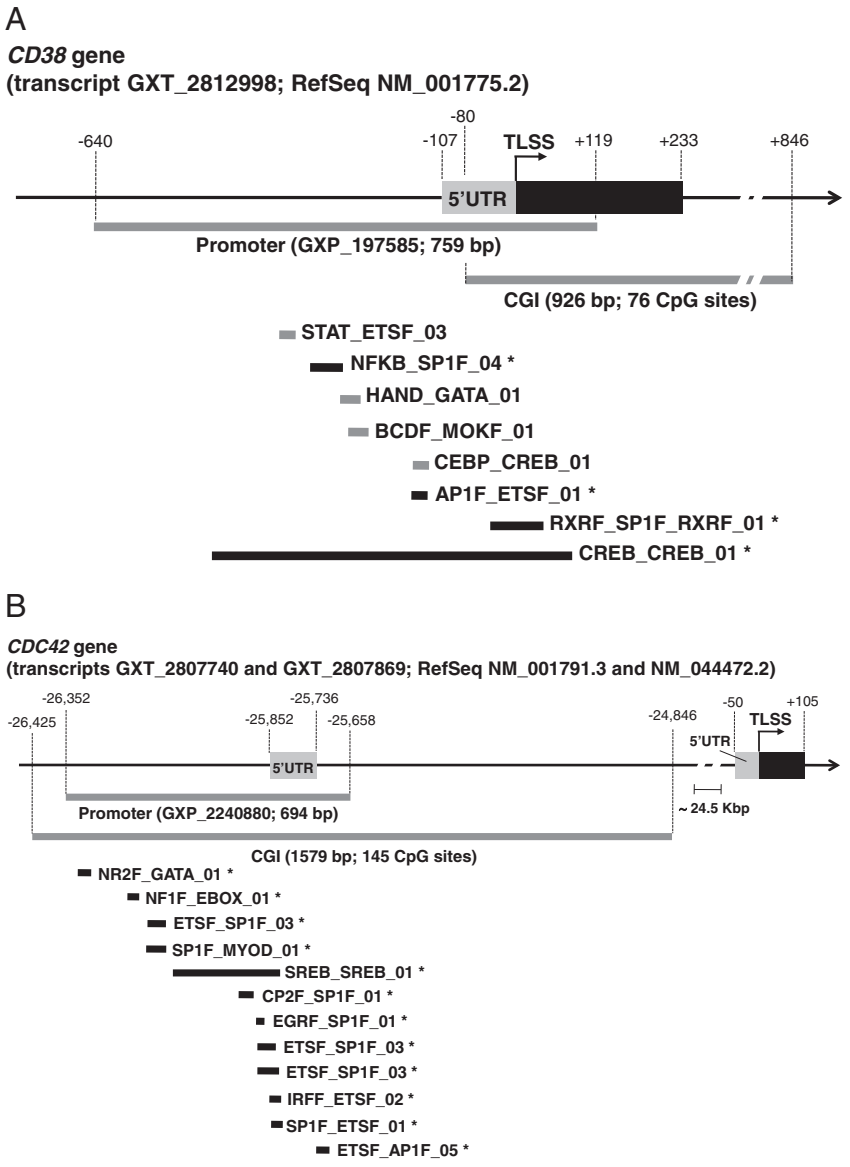


Fig. 2. Promoter region analysis of the obesity candidate gene validated transcripts. Only validated transcripts with a Reference Sequence number (NM) were selected for the promoter region analysis, and they were all identified by a GXT number within the Genomatix software. The putative promoter for each transcript was localized using the Gene2Promoter tool and identified with a GXP number. Promoter CGIs were identified using the CpG Island Explorer software and their complete localization was obtained using the NCBI Map Viewer public database. Binding sites of putative promoter TFBMs were identified using the ModelInspector tool. Below the promoter region, TFBMs covering one or multiple CpG sites are displayed with a black line and an * while those not covering a CpG site are displayed with a gray line. Delimitation of the 5′UTR, exons, putative promoters, CGIs and TFBMs were reported relatively to the first adenine (+1) from the ATG translation start site. (A) The validated transcript NM_001775.2 for the *CD38* gene had a putative promoter of 759 bp covered partly by a CGI of 926 bp, and contained 8 potential TFBMs where 4 of them were covering at least one CpG site. (B) The validated transcripts NM_001791.3 and NM_044472.2 for the *CDC42* gene had the same putative promoter of 694 bp entirely covered by a CGI of 1579 bp, and contained 12 potential TFBMs that were all covering at least one CpG site. (C) The validated transcript NM_002343.2 for the *LTF* gene had a putative promoter of 643 bp covered partly by a CGI of 561 bp, and contained 5 potentials TFBMs where 4 of them were covering at least one CpG site. (D) The validated transcript NM_002353.2 of the *TACSTD2* gene had a putative promoter of 601 bp covered partly by a CGI of 1762 bp, and contained 7 TFBMs where 3 of them where covering at least one CpG sites. bp: base pair, CD38: cluster of differentiation 38, CDC42: cell division cycle 42, CGI: CpG island, LTF: lactotransferrin, RefSeq: Reference Sequence, TACSTD2: tumour-associated calcium signal transducer 2, TFBM: transcription factor binding module, TLSS: translation start site, UTR: untranslated region.

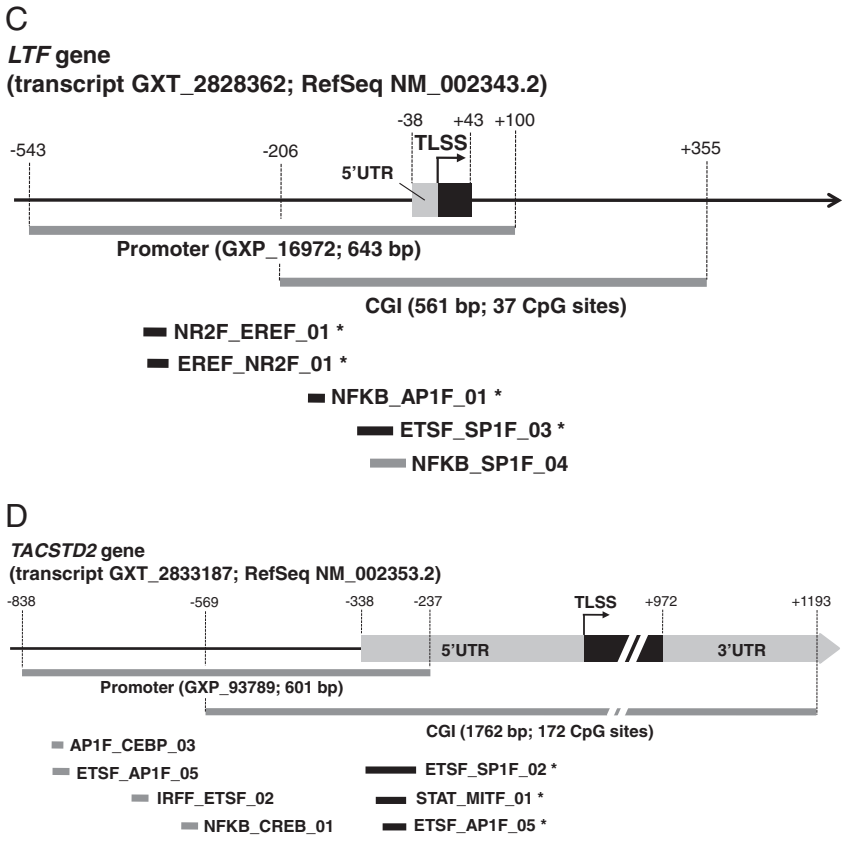


Fig. 2 (*continued*).

NM_044472.2 and NM_001791.3) that span 2308 bp (7 exons), 1530 bp (6 exons) and 2182 bp (6 exons) respectively. They all encode a functional protein of 191 aa, but the NM_001791.3 transcript differs in the 3′ region. Gene2Promoter identified a putative promoter for NM_001039802.1 transcript but no promoter CGI was present. However, a putative promoter of 694 bp covering the first 5′ untranslated region (UTR) exon (~26 kbp from the TLSS) was identified for NM_001791.3 and NM_044472.2 transcripts which is covered by a CGI of 1579 bp (Fig. 2B). ModelInspector identified 12 TFBMs in the promoter sequence and all of them had putative TFBSs that covered CpG sites. The *LTF* gene is located at the 3p21.31 locus and codes for one validated transcript (RefSeq ID: NM_002343.2) that spans 2390 bp and contains 17 exons. It encodes a functional protein of 710 aa. Gene2Promoter identified a putative promoter of 643 bp covering 543 bp before and 100 bp after the TLSS (Fig. 2C). A promoter CGI of 561 bp was observed where it covered the last 306 bp of the putative promoter and 255 bp beyond this promoter. ModelInspector identified 5 TFBMs in the promoter sequence and 4 of them had putative TFBSs that covered CpG sites. Finally, the *TACSTD2* gene is located at the 1p32 locus and codes for one validated transcript (RefSeq ID: NM_002353.2) that spans 2080 bp and contains 1 exon. This transcript encodes a functional protein of 323 aa. Gene2Promoter identified a putative promoter of 601 bp which ends 237 bp before the TLSS (Fig. 2D). A promoter CGI of 1762 bp was identified which covered the last 332 bp of the putative promoter and 1430 bp beyond this promoter. ModelInspector identified 7 TFBMs in the promoter sequence and 3 of them had putative TFBSs that covered CpG sites.

4. Discussion

This report proposes an approach for the prioritisation of a list of obesity candidate genes for which their expression may be putatively regulated by DNA methylation in promoter CGIs using gene expression data and a suite of bioinformatic tools. With strict selection criteria the gene expression profiling of peripheral blood RNA revealed 60 genes ≥2.0 fold differentially expressed between low *vs*. high BMI children and culminated in a short list of 4 target genes (*CD38*, *CDC42*, *LTF* and *TACSTD2*) putatively epigenetically regulated which can be taken forward for detailed investigation. Relying on available resources and the focus of the study the proposed workflow may be adapted in several ways to identify a greater number of genes to prioritise for further analysis.

This approach has several strengths, one being its flexibility. A greater number of prioritised genes can be obtained by customising the selection criteria, such as decreasing the fold change threshold for the selection of differentially expressed genes and/or removing the literature search step to allow the inclusion of novel genes potentially influencing obesity risk. Other advantages of this workflow rely on its ability to be applied to differential gene expression data (openly accessible in many instances) in any disease context; it begins with the previously observed differential expression, which is an asset; and it is also an inexpensive method of prioritisation precluding costly assay development and optimisation. This approach can even be applied across species, for example to interrogate data from animal models and produce a list of target genes to investigate in humans.

Although gene expression microarrays provide valuable information for deciphering the aetiopathogenesis of complex diseases we have to be aware that the differentially expressed genes are totally contingent on the quality of the expression data used initially and thus appropriate quality control and pre-processing are necessary in order to be confident that gene expression profiling data are meaningful. Even though these control steps have been done in this study other limitations can be put forth. Firstly, it is difficult to distinguish whether the differential expression observed in the study cohort are a cause or a response to different adiposity status. This is of course a

common feature of gene expression studies (Relton and Davey, 2010; Schadt et al., 2005) rather than being an issue of the *in silico* methodology itself. In this report the inclusion of the literature search may give an idea on the type of association (causal *vs*. consequence) linking the adiposity phenotypes with the candidate gene differential expression levels. Subsequent detailed analysis may not be able to delineate further the issue of reverse causation, but various strategies can be adopted to interrogate this once target loci have been identified. Nevertheless, if these prioritised loci prove to exhibit DNA methylation changes in association with an overweight/obese phenotype, analysis of epigenetic patterns in those genes may also reveal a way to identify higher-risk individuals (Relton and Davey, 2010), which is also valuable. A second limitation is that we prioritised the most likely targets by restricting the genes with $\geq$2.0 fold differential expression and selecting those associated with adiposity phenotypes which will undoubtedly overlook a large proportion of relevant loci and many potentially novel loci. The restrictive selection criteria used in this report were set on the basis of a first exploratory analysis, which can be then iterative; *i.e.* following epigenetic pattern association analyses with the first prioritised genes and adiposity phenotypes, other differentially expressed targets may be selected for further investigations. Thus, novel loci with the greatest fold differential expression or obesity-candidate genes with lower differential expression may be targeted in subsequent prioritisation processes. A third limit of this methodology regards the targeting of CpG islands localized in promoters only, which overlook other putative functional regions regulated by methylation, such as intragenic CpG islands (Deaton et al., 2011) or those localized within 2 Kb of islands (Irizarry et al., 2009). Since little is known about the functionality of these regions on gene expression regulation to date, we prioritised regions within promoter CGIs that may be bound by TFBMs, which has a logical relationship in gene expression regulation. With respect to these limitations, we are aware that the proposed approach will not identify all putatively epigenetically regulated regions associated with adiposity, although it does provide an efficient mode of prioritisation utilising several *in silico* resources.

The subjects selected for the expression profiling in this study were children aged 11–13, born prematurely and followed-up for clinical examination of cardiometabolic traits (Cooke et al., 1999, 2001). Whether the prematurity context may reveal a distinct list of differentially expressed genes between low *vs*. high BMI groups as compared to term children is possible. A recent report by Novakovic et al. (2011) has highlighted widespread changes in promoter methylation profiles in human placental tissue in response to increasing gestational age, suggesting that preterm infants may plausibly exhibit different epigenetic signatures to those born at term. Whether these differences have any bearing on subsequent phenotype can only be speculated at this stage. Apart from this particularity further epigenetic analyses in the promoter CGIs of prioritised genes may still reveal interesting associations with measures of adiposity in both children born preterm and at term. This issue may be supported by the fact that the prioritised genes were selected based on previously known associations with adiposity phenotypes. Another concern regards the use of whole blood for the identification of differentially expressed genes between children with low *vs*. high BMI. The expression profiling of this compartment cannot necessarily represent what would be seen in other tissues (tissue-specific expression) and it may also reflect a variability in blood cell population between BMI groups, as it has been seen for peripheral T cell subsets in obesity (Han et al., 2011; Svec et al., 2007). However, this cannot rule out the possibility that epigenetic mechanisms may be associated with differential expression levels observed between BMI groups in blood, that it may also reveal potential blood-based biomarkers of obesity risk, and that it may target some biological pathways implicated in obesity development. It is also possible to apply *post hoc* approaches, using data from publically available sources, to interrogate whether those genes observed to be differentially expressed show distinctly different expression signatures in B and T cells.

In conclusion, we identified 4 obesity-candidate genes putatively regulated by DNA methylation using gene expression microarray and *in silico* analysis. Further analyses exploring epigenetic patterns and adiposity associations across these genes are now warranted. Given the current status of the field and the motivation to find concrete examples of epigenetic variation associated with specific phenotypic traits and disease outcomes, the proposed method provides a viable, cost effective solution to facilitate advances in this field.

Supplementary materials related to this article can be found online at doi:10.1016/j.gene.2012.02.001.

Acknowledgments

Many thanks are expressed to the Preterm Birth Growth Study families and the associated research team. AG is funded and this study was supported by a grant from the Biotechnology and Biological Sciences Research Council (BBSRC BB/F007981/1). This research was conducted by VT during a training studentship undertaken in 2010 at Newcastle University, hosted by the laboratory of CLR and funded by the *Fonds québécois de la recherche sur la nature et les technologies* and the Institute of Nutraceuticals and Functional Foods at *Université Laval* (Québec, Canada). VT also received studentship awards from the Canadian Institute of Health Research and the *Fonds de la recherche en santé du Québec* (Québec, Canada).

References

Aksoy, P., et al., 2006a. Regulation of SIRT 1 mediated NAD dependent deacetylation: a novel role for the multifunctional enzyme CD38. Biochem. Biophys. Res. Commun. 349, 353–359.
Aksoy, P., White, T.A., Thompson, M., Chini, E.N., 2006b. Regulation of intracellular levels of NAD: a novel role for CD38. Biochem. Biophys. Res. Commun. 345, 1386–1392.
Barbosa, M.T., et al., 2007. The enzyme CD38 (a NAD glycohydrolase, EC 3.2.2.5) is necessary for the development of diet-induced obesity. FASEB J. 21, 3629–3639.
Becker, K.G., et al., 2003. PubMatrix: a tool for multiplex literature mining. BMC Bioinforma. 4, 61.
Campion, J., Milagro, F.I., Martinez, J.A., 2009. Individuality and epigenetics in obesity. Obes. Rev. 10, 383–392.
Cheng, D., Knox, C., Young, N., Stothard, P., Damaraju, S., Wishart, D.S., 2008. PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. Nucleic Acids Res. 36, W399–W405.
Chini, E.N., 2009. CD38 as a regulator of cellular NAD: a novel potential pharmacological target for metabolic conditions. Curr. Pharm. Des. 15, 57–63.
Cole, T.J., Bellizzi, M.C., Flegal, K.M., Dietz, W.H., 2000. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 320, 1240–1243.
Cooke, R.J., et al., 1999. Feeding preterm infants after hospital discharge: effect of diet on body composition. Pediatr. Res. 46, 461–464.
Cooke, R.J., Embleton, N.D., Griffin, I.J., Wells, J.C., McCormick, K.P., 2001. Feeding preterm infants after hospital discharge: growth and development at 18 months of age. Pediatr. Res. 49, 719–722.
Deaton, A.M., et al., 2011. Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res. 21, 1074–1086.
Farahani, P., et al., 2004. Obesity in BSB mice is correlated with expression of genes for iron homeostasis and leptin. Obes. Res. 12, 191–204.
Fernandez-Real, J.M., et al., 2010. Fat overload induces changes in circulating lactoferrin that are associated with postprandial lipemia and oxidative stress in severely obese subjects. Obesity. (Silver. Spring) 18, 482–488.
Ferrero, E., Saccucci, F., Malavasi, F., 1999. The human CD38 gene: polymorphism, CpG island, and linkage to the CD157 (BST-1) gene. Immunogenetics 49, 597–604.
Fornaro, M., et al., 1995. Cloning of the gene encoding Trop-2, a cell-surface glycoprotein expressed by human carcinomas. Int. J. Cancer 62, 610–618.
Frisch, M., Klocke, B., Haltmeier, M., Frech, K., 2009. LitInspector: literature and signal transduction pathway mining in PubMed abstracts. Nucleic Acids Res. 37, W135–W140.
Gentleman, R.C., et al., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80.
Godfrey, K.M., et al., 2011. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. Diabetes 60, 1528–1534.
Grant, D.J., Shi, H., Teng, C.T., 1999. Tissue and site-specific methylation correlates with expression of the mouse lactoferrin gene. J. Mol. Endocrinol. 23, 45–55.
Groom, A., et al., 2012. Postnatal growth and DNA methylation are associated with differential gene expression of the TACSTD2 gene and childhood fat mass. Diabetes 61, 391–400.

Han, S.N., Jeon, K.J., Kim, M.S., Kim, H.K., Lee, A.J., 2011. Obesity with a body mass index under 30 does not significantly impair the immune response in young adults. Nutr. Res. 31, 362–369.
Herrera, B.M., Keildson, S., Lindgren, C.M., 2011. Genetics and epigenetics of obesity. Maturitas 69, 41–49.
Hishikawa, D., et al., 2005. Identification of genes expressed differentially in subcutaneous and visceral fat of cattle, pig, and mouse. Physiol. Genomics 21, 343–350.
Hoffmann, R., Valencia, A., 2005. Implementing the iHOP concept for navigation of biomedical literature. Bioinformatics 21 (Suppl. 2), ii252–ii258.
Hong, F., Breitling, R., McEntee, C.W., Wittner, B.S., Nemhauser, J.L., Chory, J., 2006. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 22, 2825–2827.
Ibragimova, I., et al., 2010. Global reactivation of epigenetically silenced genes in prostate cancer. Cancer Prev. Res. (Phila) 3, 1084–1092.
Irizarry, R.A., et al., 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. 41, 178–186.
Jaffe, T., Schwartz, B., 2008. Leptin promotes motility and invasiveness in human colon cancer cells by activating multiple signal-transduction pathways. Int. J. Cancer 123, 2543–2556.
Janssen, I., et al., 2005. Utility of childhood BMI in the prediction of adulthood disease: comparison of national and international references. Obes. Res. 13, 1106–1115.
Jeronimo, C., Esteller, M., 2010. DNA methylation markers for prostate cancer with a stem cell twist. Cancer Prev. Res. (Phila) 3, 1053–1055.
Klingenhoff, A., Frech, K., Quandt, K., Werner, T., 1999. Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity. Bioinformatics 15, 180–186.
Kuczmarski, R.J., et al., 2000. CDC growth charts: United States. Adv. Data 1–27.
McKay, J.A., Adriaens, M.E., Ford, D., Relton, C.L., Evelo, C.T., Mathers, J.C., 2008. Bioinformatic interrogation of expression array data to identify nutritionally regulated genes potentially modulated by DNA methylation. Genes Nutr. 3, 167–171.
Moreno-Navarrete, J.M., Ortega, F.J., Ricart, W., Fernandez-Real, J.M., 2009. Lactoferrin increases (172Thr)AMPK phosphorylation and insulin-induced (p473Ser)AKT while impairing adipocyte differentiation. Int. J. Obes. (Lond) 33, 991–1000.
Novakovic, B., et al., 2011. Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors. BMC Genomics 12, 529.
Portela, A., Esteller, M., 2010. Epigenetic modifications and human disease. Nat. Biotechnol. 28, 1057–1068.
Rankinen, T., et al., 2006. The human obesity gene map: the 2005 update. Obesity (Silver. Spring) 14, 529–644.
Relton, C.L., Davey, S.G., 2010. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. PLoS Med. 7, e1000356.
Report, N.I.H., 1998. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults—the evidence report. National Institutes of Health. Obes. Res. 6 (Suppl. 2), 51S–209S.
Scarborough, P., Bhatnagar, P., Wickramasinghe, K.K., Allender, S., Foster, C., Rayner, M., 2011. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006–07 NHS costs. J. Public Health (Oxf).
Schadt, E.E., et al., 2005. An integrative genomics approach to infer causal associations between gene expression and disease. Nat. Genet. 37, 710–717.
Shaheduzzaman, S., et al., 2007. Silencing of lactotransferrin expression by methylation in prostate cancer progression. Cancer Biol. Ther. 6, 1088–1095.
Singh, A.S., Mulder, C., Twisk, J.W., van, M.W., Chinapaw, M.J., 2008. Tracking of childhood overweight into adulthood: a systematic review of the literature. Obes. Rev. 9, 474–488.
Stamatakis, E., Wardle, J., Cole, T.J., 2010a. Childhood obesity and overweight prevalence trends in England: evidence for growing socioeconomic disparities. Int. J. Obes. (Lond) 34, 41–47.
Stamatakis, E., Zaninotto, P., Falaschetti, E., Mindell, J., Head, J., 2010b. Time trends in childhood and adolescent obesity in England from 1995 to 2007 and projections of prevalence to 2015. J. Epidemiol. Community Health 64, 167–174.
Stepanow, S., et al., 2011. Allele-specific, age-dependent and BMI-associated DNA methylation of human MCHR1. PLoS One 6, e17711.
Sullivan, E.L., Grove, K.L., 2010. Metabolic imprinting in obesity. Forum Nutr. 63, 186–194.
Svec, P., et al., 2007. Do regulatory T cells contribute to Th1 skewness in obesity? Exp Clin Endocrinol Diabetes 115, 439–443.
Takai, D., Jones, P.A., 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc. Natl. Acad. Sci. U. S. A. 99, 3740–3745.
Teng, C., Gladwell, W., Raphiou, I., Liu, E., 2004. Methylation and expression of the lactoferrin gene in human tissues and cancer cells. Biometals 17, 317–323.
Usui, I., Imamura, T., Huang, J., Satoh, H., Olefsky, J.M., 2003. Cdc42 is a Rho GTPase family member that can mediate insulin signaling to glucose transport in 3T3-L1 adipocytes. J. Biol. Chem. 278, 13765–13774.
WHO, 2006. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr. (Suppl. 450), 76–85.
Yagi, M., et al., 2008. Lactoferrin suppress the adipogenic differentiation of MC3T3-G2/PA6 cells. J. Oral Sci. 50, 419–425.