Comprehensive Assessment of the Role of DNA Methylation in Obesity and Type 2 Diabetes

Alexander Werner Drong

Wellcome Trust Centre for Human Genetics

University of Oxford

Christ Church, Trinity Term 2014

Thesis submitted to the University of Oxford in candidature for the degree Doctor of Philosophy

Memorandum

The work in this thesis is the original work of the author. Analyses were carried out under the supervision of Dr Cecilia Lindgren and Prof Mark McCarthy. Funding was provided by the Wellcome Trust.

The DNA methylation data from adipose tissue in Chapter 3 were generated as part of the MolOBB study by Epigenomics AG (Berlin, Germany) using differential methylation hybridisation. Normalised and log2-transformed Gene Expression data were provided by Dr George Nicholson (University of Oxford). Replication data from the mUTHER study was provided by Dr Eshwar Meduri (Wellcome Trust Sanger Institute). The resulting manuscript is attached as an appendix.

The DNA methylation data from whole blood in Chapter 4/5 were generated as part of the EpiMigrant study by the Oxford Genomics Centre. Annotations of SNPs within methylation probes based on a South Asian reference panel, permutated and spiked methylation intensities, estimated white blood cell counts and parallelized scripts used for association testing were provided by Dr Benjamin Lehne (Imperial College London).

The gene expression results in Chapter 5 were meta-analysed from KORA (Simone Wahl, Helmholtz Institute Munich) and LOLIPOP (Dr Marie Loh, Imperial College London) data by Simone Wahl. The KORA association results used in the meta-analysis of clinical traits were generated by Simone Wahl. The Mendelian Randomisation experiments were performed by Dr Benjamin Lehne.

This thesis is approximately 43,000 words.

Comprehensive Assessment of the Role of DNA Methylation in Obesity and Type 2 Diabetes

Alexander Werner Drong, Christ Church

Submitted in candidature for the degree of Doctor of Philosophy

Trinity Term 2014

Abstract

Obesity and type 2 diabetes (T2D) are major risk factors for cardiovascular and other diseases and are currently undergoing an increase in global prevalence. The work presented in my thesis addresses the role epigenetics, specifically DNA methylation, plays in the susceptibility to obesity and T2D and deals with methodological issues in the analysis of DNA methylation data.

I first combined epigenome-wide DNA methylation data across 38 adipose tissue samples with corresponding SNP and mRNA data for the same subjects. At 5% false discovery rate (FDR), methylation of 149 regions associated with at least one *cis*-SNP. When 19 of the 149 regions were tested for association in an additional 181 independent samples, five regions replicated. These results indicate a genetic influence on DNA methylation in adipose tissue.

I then analysed 90 epigenome-wide methylation samples taken from 15 South Asian controls and 30 T2D cases participating in the LOLIPOP study at two time points ~7 years apart. I found global differences at both follow-up and baseline between the normal glucose tolerant and T2D groups, as well as strong differences with aging. I further used the main EpiMigrant data from 2,687 individuals, with 36 samples measured in duplicate to assess approaches to quality control, data normalisation and batch correction through control probe adjustment. A null hypothesis for epigenome-wide association studies (EWAS) by permutation testing and I investigated the effects of correlation between individual methylation markers.

Using the developed methods, I carried out an EWAS of body mass index (BMI) with subsequent meta-analysis amongst 10,261 individuals of European and South Asian ancestry. DNA methylation markers at 187 genetic loci were associated with BMI. Mendelian randomisation experiments suggested that association of DNA methylation with BMI is the consequence of BMI.

Lastly, I tested haplotypes of 85 SNPs currently known to be associated with T2D and 118 SNPs associated with obesity traits for an enrichment of CpG creating or abrogating SNPs and found that 9 T2D and 23 obesity SNPs showed a significant difference in CpG count between the SNP alleles as established by permutation testing. Amongst these is *FTO*, a locus which has been previously been shown to have a haplotype-specific methylation effect.

My work provides novel insights into the role of DNA methylation in metabolic diseases. The methods that I developed to robustly detect association are flexible and scalable and will further be useful for larger, future EWAS.

Acknowledgements

First and foremost I would like to thank my supervisors Cecilia Lindgren and Mark McCarthy. My thanks not only extend to their scientific support and guidance through the course of this thesis, but also for being an inspiration ever since I met both of them during my summer project in 2007. By introducing me to several collaborators and sending me to lab exchanges around the world, they truly made this DPhil experience unique and took care of my personal and professional development. I would also like to thank the Wellcome Trust for the generous funding over the course of four years.

I would also like to thank Dr John Chambers, and his team (Benni, Marie and Weihua) who also took me under their wings and hosting me as a visiting researcher for three years to work on the EpiMigrant samples. I would also like to thank my mentors at Oxford, Dr Mattias Rantalainen and Dr George Nicholson, who helped me with any statistical problems.

I would also like to thank my college, Christ Church, with all my friends in both the JCR and GCR. Going into my 9th year there and two degrees later, I cannot think of a more beautiful place to have spent my youth and I am proud to still have my home base there. My friends at my college and all over Oxford are way to many to mention by name, but I'll try: Tristan x2, Kate, Meelis, Jon Cleland, Chris Lander, Narin, Annemari, Sophie, Jon Storey, Shari, Sonya, Alex B, Evan, Michael, Jian, Sylvestre, Raam, Christiaan, Pete, Alex Lunt, Yuepeng, Dawnson, Andi Schmidt, Dragos, Eugene Ong, Tim, Justin, Dom, Helen, Jake, Martin Grossel and the whole PresCom group. My apologies if I forgot anyone.

I would like to also particularly thank my friend Paul Robertson, a talented DPhil student on the Wellcome Trust programme who started in the same year as me, but sadly passed away during our final year of study. I miss you deeply buddy.

Lastly, I would like to thank my extended family (Werner, Heidi, Tina, Sepp, Dorchen, Gerd, Ingrid, Michael, Angela, Benni, Florian, Peter, Barabara, Helmut, Elke, Svenja, Christina), friends of the family (Tihomir, Anja, Meik, Matteo and Emilia) and friends from Hannover (The "Five" Kings: Chris, Thomas, Henne, Lennart, Marius, Moritz, Paul, Timm, Timo, Lukas). They have always provided guidance and support in all my years of study. And of course Jess, who gave me emotional support and peace of mind when the thesis write-up got stressful. Thank you!

List of common abbreviations

27k array	Illumina Infinium HumanMethylation27 BeadChip
450k array	Illumina HumanMethylation450 BeadChip
ADA	American Diabetes Association
ASM	allele-specific DNA methylation
A^{vy}	agouti viable yellow
BMI	body mass index
BMIQ	Beta Mixture Quantile Normalisation
BP	blood pressure
ChIP	chromatin immunoprecipitation
CI	confidence interval
CNV	copy number variant
CT	computed tomography
DEXA	dual energy x-ray absorptiometry
DNMT	DNA methyltransferase
DMH	differential methylation hybridisation
DMP	differentially methylated position
DMR	differentially methylated region
EWAS	epigenome-wide association study
FDR	false discovery rate

GC	genomic control
GDM	gestational diabetes mellitus
GRS	genomic risk score
GSEA	gene set enrichment analysis
GWAS	genome-wide association study
H^2	broad sense heritability
h^2	narrow-sense heritability
HAT	histone acyl transferases
HbA1c	glycated haemoglobin
HDAC	histone deacetylases
HDL	high-density lipoprotein
HSM	haplotype-specific methylation
IHEC	International Human Epigenome Consortium
Illumina	Illumina Control Probe normalisation
IQR	interquartile range
LCL	lymphoblastoid cell line
LD	linkage disequilibrium
LDL	low-density lipoprotein
LOLIPOP	London Life Sciences Population
meQTL	methylation quantitative trait locus

MetSyn	metabolic syndrome
MHC	major histocompatibility complex
miRNA	micro-RNA
MODY	maturity-onset diabetes of the young
MR	Mendelian Randomisation
MRI	magnetic resonance imaging
NAFLD	non-alcoholic fatty liver disease
ND	neonatal diabetes
NGT	normal glucose tolerant
OGTT	oral glucose tolerance test
PBC	peak-based correction
PBL	peripheral blood leucocytes
PBS	Portable Batch System
PC	principal components
PCA	principal component analysis
PCR	polymerase chain reaction
QC	quality control
QN(B)	quantile normalisation of beta values
QN(I)	quantile normalisation on intensity values
RMA	robust multi-array average

RNAi	RNA interference
SAT	subcutaneous adipose tissue
SD	standard deviation
siRNA	small-interfering RNA
SNP	single nucleotide polymorphism
SVD	singular value decomposition
SWAN	Subset within-array normalisation
T-T2D	treated T2D
T1D	type 1 diabetes
T2D	type 2 diabetes
TC	total cholesterol
TG	triglycerides
Tost Pipeline	Subset quantile normalization
TSS	transcription start site
U-T2D	untreated T2D
VAT	visceral adipose tissue
WC	waist circumference
WHO	World Health Organization
WHR	waist-hip-ratio

Table of Contents

Chapter 1 Introduction

1.1 Global Burden of T2D and Obesity

Type 2 diabetes (T2D) and obesity are diseases that are currently undergoing a global increase in prevalence and present major public health and economic problems. The most recent global study from 2013 estimates that there are 382 million people (8.3% of all adults) with T2D and projects this number to increase by 55% to 592 million by 2035.[1] The annual growth of thus 2.7% is about 1.7 times the growth in the global population.[2] A large proportion (80%) of the people with T2D live in low- to mid-income countries where the medical systems are least likely to cope: of the overwhelming estimated $548 billion spent on healthcare costs of T2D worldwide, only 1% are accounted for by South-East Asia and Africa.[1] The proportion of global deaths attributable to T2D worldwide is estimated to be ~7%.[3] In total, 5.1 million deaths were caused in total by all forms of diabetes.[1]

Obesity, which is an important risk factor for T2D, is also projected to globally increase in prevalence. In 2013, 29.8% of men and 38.0% of women worldwide were estimated to be overweight or obese. Of these, one third are obese. Obesity-related diseases, such as osteoarthritis and many cancers add an estimated $48–66 billion a year to healthcare costs in the USA.[4] In 2010, 3.4 million deaths were caused by obesity as defined above and 3.9 of years of life lost on average.[5]

1.2 Pathophysiology of T2D and Obesity

1.2.1 The different types of diabetes

T2D is a metabolic disease, which is characterised by sustained elevated plasma glucose levels (hyperglycaemia).[6] This condition occurs when insulin secretion from the pancreatic islets does not sufficiently counteract increasing insensitivity to the

action of circulating insulin on its target tissues.[1] Various forms of diabetes (type 1 diabetes [T1D][7], T2D and gestational diabetes mellitus [GDM][8]) are recognised, as well as more rare, monogenic and syndromic forms including MODY (maturity-onset diabetes of the young).[9] Of these, T2D is the most common, accounting for 85%-95%[1] of the diabetes incidence, followed by T1D with 5%-10%[1], leaving 1%-2% for the remaining rarer forms of diabetes.[1] In **Table 1.1**, the characteristics of each diabetes subclass are described.

Table 1.1: The different subtypes of Diabetes

Type	Onset	Characteristics	Pathophysiology	Prevalence
Type 1 Diabetes	Sudden onset in children or adolescents	Most severe form, needs daily insulin treatment. Episodes of Hypoglycaemia are common	Autoimmune reaction against beta cells	5-10%
Gestational Diabetes	In women during pregnancy, typically around 24 weeks	Foetal macrosomia, increased risk	Placental hormones blocking insulin action	1-2%
Type 2 Diabetes	Usually in adulthood, patients may not have symptoms before diagnosis	No episodes of hypoglycaemia, can be prevented by healthy diet and exercise	Insulin resistance of glucose dependent tissues with subsequent decreased insulin secretion	85-95%
Maturity Onset Diabetes of the Young	Typically in adolescents	Treatment can be optimised if diagnosed correctly	Rare coding mutations	1-2%

1.2.1.1 Clinical Characteristics of T2D

Different diagnostic criteria for T2D were set out by World Health Organization (WHO)[10] and American Diabetes Association (ADA)[11] based on alternative measures of glycaemia: plasma glucose levels (fasting glucose >7 mmol/L or non-fasting >11.1

mmol/L), two hour plasma glucose concentration > 11.1 mmol/L after an oral glucose tolerance test (OGTT) or glycated haemoglobin (HbA1c) > 48 mmol/mol.[10-12] However, if there are no symptoms of diabetes present, the diagnosis should be confirmed by repeat testing. Recent work has also shown that there are ethnicity specific differences in HbA1c independent of glucose[13], which need to be taken into account.

1.2.1.2 General Symptoms and Complications

Typical symptoms of diabetes can vary from patient to patient and depending on the severity of the condition, but commonly include polydipsia (abnormal thirst) and dry mouth, polyuria (excessive urination), lack of energy and tiredness, hunger, unexplained weight loss, slow-healing wounds, recurrent infections and blurred vision.[1] As the symptoms can often be moderate, T2D is often referred to as a mild form of diabetes that (unlike T1D) often persists unnoticed in patients for many years. However, the unawareness and thus poor treatment of glucose imbalance often results in prolonged damage of the body and long-term complications leading to the increased mortality in T2D patients described above.[14] Diabetic complications[15] include cardiovascular events[16], such as angina, myocardial infarction, strokes, peripheral artery disease, as well as microvascular events including kidney failure (nephropathy)[17], visual disability (retinopathy).[1, 18] Long-term studies in the UK[19] have shown that hyperglycaemia itself is one of the major causes of cardiovascular complications, highlighting the need for good glycaemic control.[19, 20] Together with excessively high blood pressure, sustained levels of elevated blood glucose can cause nerve damage (neuropathy).[21] The adverse effects are particularly pronounced on the extremities, where a loss of feeling can lead to ulcerations and infections (diabetic foot).[21] Diabetic foot is the most common form of amputation in diabetes patients, with 15% of people with diabetes developing a foot problem. Of these, 10% will require amputation.[22]

1.2.2 Obesity

Obesity is a condition where excess body fat in the form of triglycerides has accumulated in a patient's body to the extent that it has a negative effect on the patient's health.[23] It is typically caused by an imbalance between caloric intake and energy expenditure.[23] The condition is one of the most common risk factors for T2D, but also harbours risk for other diseases. In developed countries, obesity is often viewed as a character flaw and obese people regarded as unattractive and often discriminated against. Thus, social stigmatisation is another problem beside the adverse health effects.[23, 24] Despite the fact that obesity has been known to be harmful, the global impact it has, especially in developing countries, has only been realized within the last 30-40 years. While underweight and malnutrition always overshadowed obesity as nutritional health problems[23], obesity now presents a greater public health issue than these conditions.[5]

1.2.2.1 Clinical Characteristics of Obesity

By WHO conventions, a person is classified as overweight at a BMI higher than 25 kg/m^2, obese at BMI > 30 kg/m^2, severely obese at BMI > 35 $kg/m^2$and morbidly obese at BMI > 40 kg/m^2.[25] A common criticism of using BMI as an obesity phenotype is that while it is easily and non-invasively measured in the clinic, it gives incomplete information about the physical state of a person, since it does not distinguish between lean and fat body mass and does not capture the distribution of adipose tissue across the body. Moreover, there is age bias as BMI is cannot be used as a reliable measure of adiposity in children or adolescents[26] as well as elderly people.[27] Measures of central obesity (waist circumference [WC]) and fat distribution (waist-hip-ratio [WHR]), are often also considered, and correlate well with BMI (R^2 with BMI ~0.9 and ~0.6, respectively).[28] These are particularly interesting to investigate, since different fat depots contribute differently to disease risk.[29] Increasingly, more sophisticated measurements such as per cent body fat/fat mass

are used, such as those measured by bioimpedance, dual energy x-ray absorptiometry (DEXA), computed tomography (CT) or magnetic resonance imaging (MRI).[23] As is the case with measuring error in BMI[30], however, noise in machine measurements, as well as lower sample sizes due to cost, can affect power in the study of adiposity as a continuous trait.[31]

1.2.2.2 Clinical Consequences of Obesity

Obesity can reduce life expectancy by up to ten years on average in cases above the "severe obesity" limit of 35 kg/m^2.[23] This morbidity can be either explained by a direct causality of obesity to the disease, or the disease sharing a common causality with obesity, such as diet and lifestyle. Moreover, the mortality of obesity increases with age.[32] In general, the diseases caused can be attributed to two factors: The first one is increased body mass, leading to poor mobility, osteoarthritis, gout, back pain, asthma and psychiatric diseases such as depression.[23] The second one is due to metabolic imbalances, which, amongst others, can lead to diabetes, cancer, cardiovascular diseases and non-alcoholic fatty liver disease (NAFLD).[4]

1.2.3 Metabolic Syndrome and T2D

There is growing interest in the role of epigenetic factors in predisposition to common and metabolic diseases. Metabolic Syndrome (MetSyn) is defined as a cluster of risk factors for T2D and heart diseases.[6, 33] The estimated 20-25% of the adult world population with MetSyn have a threefold greater risk of a heart attack or stroke, and a fivefold increased risk of T2D compared to people with the syndrome.[6]

The diagnostic criteria outlined by the International Diabetes Foundation[6] require central obesity (with ethnicity-specific values of WC, or assumed central obesity if the patient's BMI exceeds 30 kg/m^2) and two of the following criteria:

- Triglycerides ≥ 150 mg/dL (or treatment for raised triglycerides)

- HDL cholesterol < 40 mg/dL in males or < 50 mg/dL in females (or treatment for reduced HDL cholesterol)
- Systolic blood pressure (BP) ≥ 130 or diastolic BP ≥ 85 mm Hg (or treatment for hypertension)
- Fasting plasma glucose ≥ 100 mg/dL or previously diagnosed T2D

The presence of metabolic syndrome together with impaired glucose tolerance or impaired fasting glucose doubles the risk of developing T2D.[34] Therefore, it has been suggested that prediabetes and MetSyn denote the same disorder with different biomarkers.[34]

1.3 Familial Aggregation: Genetics, Epigenetics and the Environment

There is substantial evidence that both T2D and obesity traits cluster within families provided by studies of populations, families and twins.[35-37] Despite the late-onset of T2D, and the variation of obesity traits over time, it has been shown that parental T2D or obesity is an important risk factor for these diseases: a study following a cohort for 40 years found that the odds ratios for developing abnormal glucose tolerance for individuals with paternal or maternal diabetes were ~3.5[35] as defined by fasting glucose levels. The sibling recurrence risk ratio for T2D is of a similar degree.[38] The genetic component of the risk is measured by the narrow-sense heritability (h^2), which captures additive effects of parental alleles on phenotypic variance in their offspring by comparing monozygotic to dizygotic twin pairs.[39]. The broad sense heritability (H^2), on the other hand, measures the total genetic contribution to the variance of a trait.[39] Commonly used estimates of h^2 can vary and may be inflated, but common current estimates are ~25% for type 2 diabetes[36] and between 40% and 70% for obesity-related traits.[37] This high apparent heritability had sparked the initial interest in genetics research in these diseases.

Convincing evidence for a genetic effect in T2D and obesity was given by studies that seek to control for the effects of either genetics or environment.[40, 41] For instance a study of 311 twin pairs reared apart and 308 twin paired reared together showed that the intrapair correlation coefficients of BMI were 0.70 for men and 0.66 for women, which was only slightly lower than those for twins reared together.[42]

Similarly, a study of 12 twin pairs undergoing overfeeding in controlled environments consistently suggested an influence of non-environmental factors and high heritability.[41] In the study, 12 pairs of young adult male monozygotic twins were overfed by 1000 kcal per day, 6 days a week, during a 100-day period, resulting in a mean weight gain of 8.1 kg.[41] For all of body weight, percentage of fat, fat mass, and estimated subcutaneous fat, there was three times more variance among pairs than within pairs. The twin similarity was even more evident for fat distribution regional fat, with about six times as much variance among pairs as within pairs.[41] These results underlined that genetics, together with the environment, play an important role in energy intake and expenditure. As monozygotic twins share the same maternal environment as well as genetics, the results could possibly also be epigenetic

Familial aggregation, however, does not directly equate to a genetic effect. The chance of developing T2D or obesity depends on exposures to environmental or metabolic stress, as well as the genetic constitution to respond adequately. In the extreme case, the familial aggregation of T2D and obesity may only reflect effects of a shared family environment, as there are clear relationships to lifestyle factors.[20] The recent dramatic increase in both T2D and obesity incidence over the past few decades[1] means that impact of dramatic changes in lifestyle and environment must be the primary cause, as genetic selection for T2D and obesity susceptibly variants could not occur in such a short time.[43] These environmental factors may tend to cluster within families[44], but are unlikely to fully explain the familial aggregation.

Psychological and social factors (such as socioeconomic status, dietary preferences, food availability) are an important risk factors of T2D and obesity, along with the environmental effect of sedentary lifestyle with less physical activities and increased energy intake.[45] One such explanation is given by the "thrifty genotype hypothesis", which states that genes promoting storage of fat have been evolutionary advantaged in times of low food availability and overweight individuals would be more likely to survive during times of famine.[43] Under steady food supplies, the human tendency to accumulate fat has led to negative health effects[43]. This theory would also apply for T2D, as obesity, age, physical inactivity and diet are the most well-characterised risk factors.[45] However, apart from the social and lifestyle environment[45, 46], there is also an effect of the more immediate chemical and biological environment, which for example includes infections and differences in gut flora.[47]

A more recent suggestion is that familial aggregation can also reflect epigenetic processes that generate patterns of intrafamilial and transgenerational resemblance.[48] For example, early life events can increase the sibling recurrence of long-term effects of on adult risk of T2D and obesity[38]. It has been suggested that mothers with diabetes during pregnancy potentially to lead to the propagation of T2D risk into successive generations through non-genetic means.[49-51] The foetal origins (also known as the "thrifty phenotype" or Barker-Hales) hypothesis[52] links events in early life such as an adverse uterine environment to the increased risk of T2D in later life. The metabolic priming likely to involves epigenetic processes, which together with genetic inheritance influence disease risk.[49]. In the following sections I will explain further the insights gained from genetic and epigenetic studies of T2D and obesity.

1.4 Genetics of T2D and obesity

Over 150 genetic loci have been robustly associated with monogenic, syndromic or multifactorial forms of obesity or (non-autoimmune) diabetes.[53] These loci, separately

and in aggregate, provide an opportunity to assess the genetic architecture and pathophysiology of these traits.

1.4.1 Animal Models

The study of the genetics of obesity and T2D was first supported by extensive animal studies. The first success was the positional cloning of the leptin gene in mice by Zhang et al. in 1994.[54] The mouse model, Lep^{ob} as well as the Zucker strains rely on the mutations in the leptin gene or receptors to induce T2D via obesity.[55] When leptin-knockout mice are injected with synthetic leptin, they return to normal body weight, thus proving the role of leptin in the accumulation of body fat (see above).[56] Subsequent research also showed that targeted disruption of the melanocortin 4 receptor (*MC4R*) gene results in obesity in mice.[57] More recent mouse models with a polygenic background include the New Zealand obese (NZO) and Otsula Long Evans Tokushima Fatty rat (OLETF), which are widely used for studying diabetes treatment and dependence of risk factors.[55]. Generating new animal models for newly found associations in humans is often used to elucidate the biological mechanisms behind the associations.

1.4.2 Monogenic diabetes

Monogenic and syndromic forms of diabetes and obesity are caused by rare, highly penetrant alleles. Family-based linkage studies were able to identify rare monogenic genes forms of diabetes, such as through the MODY families.[9] While most cases are rare coding mutations in either the hepatocyte nuclear factor-1A (*HNF1A*) or glucokinase (*GCK*) genes, at least five genes were found: *HNF4A, PDX1, HNF1B, NEUROD1, INS.*[58] The study of MODY has most importantly yielded treatment optimization, as well as a better understanding of glucose homeostasis.[59]

Another rare, non-autoimmune disease of interest is neonatal diabetes (ND), which occurs within the first week of birth and can either be permanent of transient.[60, 61] The

latter has a duration of typically 10 weeks, but increases risk of developing T2D later in life[61]. Studies have found various loci associated with ND: *KCNJ11*[62] and *ABCC8*[63, 64] which encode for subunits of an ion channel key for insulin release and accounts for cases of permanent or transient ND.[62, 65] Other ND loci include *INS*[65], *HNF1B*[66] and imprinted genes *PLAGL1* and *HYMAI*. At these loci, imprinting, which is a process of silencing genes and establishing a parent-of-origin effect[67], might be controlled by *ZFP57* mutations.[61]

1.4.2.1 Monogenic Obesity

Analogously, certain cases of extreme childhood obesity can be explained by monogenic mutations affecting leptin (*LEP*), leptin receptor (*LEPR*), *MC4R* or *POMC.*[68, 69] The identification of leptin was very significant, as it sparked the study of pharmacological therapy of obesity and led to the understanding of the regulation of body weight.[70] Mutations in genes in the leptinergic-melanocortinergic system can lead to excess body fat[71], which can be treated with recombinant leptin in children.[72] Further, the identification of the mutations underlying syndromic forms of obesity[73-75], including Bardet-Biedl, have discovered that these pleiotropic diseases result from defects in ciliary development and function.[76] However, as monogenic forms are caused by severe and rare single-gene mutations, they cannot account for the current epidemic of obesity and T2D.[77]

1.4.3 Linkage and Candidate Gene Studies

For common forms of diabetes and obesity, a more appropriate model is that of multiple genetic and environmental factors, which in combination produce a measurable difference.[56] Initially, small-scale candidate gene studies were undertaken using obese/T2D cases and normal controls, but as most obesity/T2D genes only have a moderate to low effect, the early studies where underpowered and relatively unsuccessful[78]. With the realisation of sample size came a few successful studies that found associations with variants in the genes encoding key therapeutic

targets for T2D such as the peroxisome proliferator-activated receptor gamma (*PPARG*) and the islet K_{ATP} channel (*KCNJ11*).[62, 79] Equivalently, for obesity this relates to variants in the *MC4R* gene[80] previously implied by family-based linkage studies, which make use of family pedigrees to map genetic loci.[81]

1.4.4 Genome-wide association studies

As early attempts to apply the linkage approaches based on families to T2D and common obesity failed, Risch and Merikangas[82] highlighted the merits of association analysis for the detection of the low penetrance alleles most likely to be relevant to common disease. However, only a decade later the density of single nucleotide polymorphism (SNP) markers available made genome-wide association studies (GWAS) feasible[83], which has led to an acceleration in the number of loci now known to be associated with common forms of T2D and obesity, as well as with closely-related continuous traits, such as fasting glucose, BMI and fat distribution.[83-87]

These studies have focused on the detection of signals attributable to common variants (minor allele frequency [MAF] above 5%) by using genotyping microarrays. There have been many GWAS looking at the genetic factors affecting T2D susceptibility. These include European[84, 85, 88-90], east Asian[91-95], south Asian[96-98], Mexican/Mexican American[98] and African American GWAS[99]. Most recently, a meta-analysis between populations of multiple ethnicities was performed replicating 69 established T2D loci and identifying seven new susceptibility loci, leading to a total of 76 robustly associated loci (**Figure 1.1, Table 1.2**).

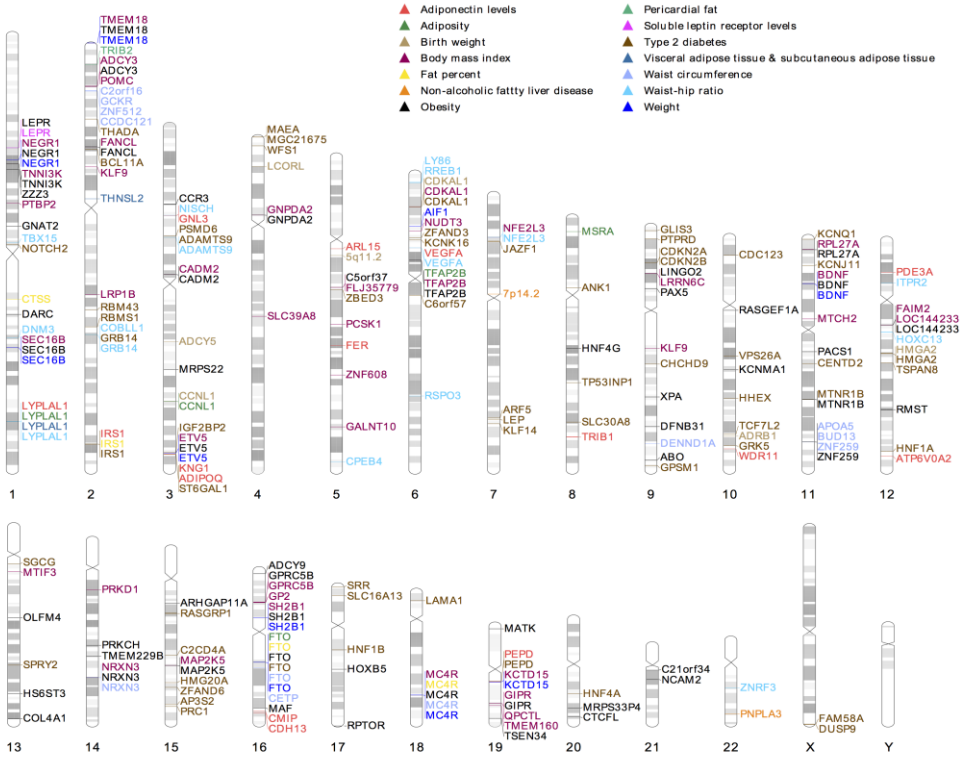


Figure 1.1: Genome-wide significant associations T2D and obesity-related traits loci. Karyogram of the human chromosomes showing approximate chromosomal positions of published genome wide significant ($p<5x10^{-8}$) GWAS hits. All the data were obtained from the National Institute of Genome Research (NHGRI) GWAS catalogue (www.genome.gov), where gene names are estimated as the nearest gene. Associations with traits are labelled with the colours outlined in the insert.

Table 1.2: Known genetic loci for T2D

SNP	Locus	Chr	Position (Mb)	Reference
rs10923931	*NOTCH2*	1	120.3	90
rs340874	*PROX1*	1	212.2	85
rs780094	*GCKR*	2	27.6	85
rs11899863, rs7578597	*THADA*	2	43.5	90
rs243088, rs243021	*BCL11A*	2	60.4	89
rs7593730	*RBMS1*	2	160.9	93
rs3923113,rs13389219	*GRB14*	2	165.2	96, 89
rs7578326,rs2943641	*IRS1*	2	226.7	90
rs13081389,rs1801282	*PPARG*	3	12.3	90, 95
rs7612463	*UBE2E2*	3	23.3	95

rs831571	*PSMD6*	3	64.0	[91]
rs6795735,rs4607103	*ADAMTS9*	3	64.7	[90]
rs11708067	*ADCY5*	3	124.5	[85]
rs1470579	*IGF2BP2*	3	187.0	[90]
rs16861329	*ST6GAL1*	3	188.1	[96]
rs10010131, rs1801214	*WFS1*	4	6.3	[90]
rs459193	*ANKRD55*	5	55.8	[89]
rs4457053	*ZBED3*	5	76.5	[90]
rs7754840,rs10440833	*CDKAL1*	6	20.8	[90]
rs9470794	*ZFAND3*	6	38.2	[91]
rs1535500	*KCNK16*	6	39.4	[91]
rs17168486, rs6960043, rs2191349	*DGKB*	7	14.9	[89]
rs849134	*JAZF1*	7	28.2	[90]
rs4607517	*GCK*	7	44.2	[85]
rs6467136	*GCC1/PAXA4*	7	127.0	[91]
rs972283	*KLF14*	7	130.1	[90]
rs516946	*ANK1*	8	41.6	[89]
rs896854	*TP53INP1*	8	96.0	[90]
rs13266634, rs3802177	*SLC30A8*	8	118.3	[90]
rs7041847	*GLIS3*	9	4.3	[91]
rs17584499	*PTPRD*	9	8.9	[100]
rs944801,rs10965250	*CDKN2A/B*	9	22.0	[89, 90]
rs10811661	*CDC123/CAMK1D*	9	22.1	[89, 90]
rs13292136	*CHCHD9/TLE4*	9	81.1	[90]
rs2796441	*TLE1*	9	83.5	[89]
rs12779790	*CDC123/CAMK1D*	10	12.4	[90]
rs1802295	*VPS26A*	10	70.6	[96]
rs12571751	*ZMIZ1*	10	80.6	[89]
rs1111875, rs5015480	*HHEX/IDE*	10	94.5	[90]
rs7903146	*TCF7L2*	10	114.7	[90, 101]
rs2334499	*DUSP8*	11	1.7	[88]
rs231362,rs163184,rs231361	*KCNQ1*	11	2.6	[90]
rs5215	*KCNJ11*	11	17.4	[90]
rs1552224	*ARAP1/CENTD2*	11	72.1	[90]
rs1387153, rs10830963	*MTNR1B*	11	92.3	[90]
rs11063069	*CCND2*	12	4.2	[89]
rs10842994	*KLHDC5*	12	27.9	[89]
rs1531343	*HMGA2*	12	64.5	[90]

rs4760790	*TSPAN8/LGR5*	12	69.9	[90]
rs7957197	*HNF1A/TCF1*	12	119.9	[90]
rs1359790	*SPRY2*	13	79.6	[102]
rs7163757	*C2CD4A*	15	60.2	[95]
rs7178572,rs7177055	*HMG20A*	15	75.5	[89, 96]
rs11634397	*ZFAND6*	15	78.2	[90]
rs2028299	*AP3S2*	15	88.2	[96]
rs8042680	*PRC1*	15	89.3	[90]
rs11642841	*FTO*	16	52.4	[90]
rs7202877	*BCAR1*	16	73.8	[89]
rs391300	*SRR*	17	2.2	[100]
rs4430796	*HNF1B/TCF2*	17	33.2	[90]
rs12970134, rs11873305	*MC4R*	18	56.0	[89]
rs10401969	*CILP2*	19	19.3	[89]
rs3786897	*PEPD*	19	38.6	[91]
rs8108269	*GIPR*	19	50.9	[89]
rs6017317	*FITM2/R3HDML/HNF4A*	20	42.4	[91]
rs4812829	*HNF4A*	20	42.4	[96]
rs5945326	*DUSP9*	23	152.6	[90]

Chr: Chromosome

Similarly, the genetics of obesity have been studied in many GWAS and genome-wide meta-analyses using different measures of body adiposity including BMI for overall adiposity[86, 103-107], body fat percentage[108], WC and WHR for central fat distribution.[87, 109, 110] In these GWAS, 32 loci associated with BMI, two novel loci associated with body fat percentage, 14 novel loci associated with WHR (adjusted for BMI, WHRadjBMI) with stronger effect in women, eight new loci associated with extreme and early-onset obesity, one novel loci associated with visceral adipose tissue (VAT) in women, five loci associated with NAFLD and one new loci associated with pericardial fat were identified (**Figure 1.1, Table 1.3**). Recent efforts are characterized by the integration and meta-analysis of data from multiple cohorts, with the largest ongoing studies for BMI including data from over 330,000 individuals (GIANT unpublished). Other studies aim at exploring the genetic basis of trait variability broader range of ethnic groups[91, 96, 111, 112], expanding on knowledge from earlier European studies.

Table 1.3: Known genetic loci and SNPs for T2D

SNP	Locus	Chr	Position (Mb)	Phenotype	Reference
rs2815752	*NEGR1*	1	72.6	BMI	86
rs1040070, rs1514175	*TNNI3K*	1	74.8	BMI, childhood obesity	86, 113
rs17381664	*ZZZ3*	1	77.8	class 2 obesity	114
rs1555543	*PTBP2*	1	96.7	BMI	86
rs17024258	*GNAT2*	1	109.9	class 1 obesity	114
rs984222	*TBX15-WARS2*	1	119.3	WHR,adjBMI	87
rs1011731	*DNM3-PIGC*	1	170.6	WHR,adjBMI	87
rs516636, rs574367,rs543874, rs10913469	*SEC16B*	1	176.1-176.2	BMI, childhood obesity	86, 109, 111-113
rs2605100, rs11118316, rs4846567	*LYPLAL1*	1	217.7-217.8	WHR,adjBMI, VATSAT	87, 115
rs2867125, rs4854344	*TMEM18*	2	0.61-0.62	BMI, Extreme obesity, childhood obesity	86, 113
rs10198628	*TRIB2*	2	12.9	Pericardial Fat	115
rs6545814	*ADCY3-DNAJC27*	2	25.0	BMI	112
rs6752378-rs713586	*POMC-RBJ*	2	25.0	BMI, childhood obesity	86, 113
rs12617233, rs887912	*FANCL-FLJ30838*	2	58.9-59.2	BMI	86, 116
rs1659258	*THNSL2*	2	88.4	VAT in women	115
rs2890652	*LRP1B*	2	142.7	BMI	86
rs10195252	*GRB14*	2	165.2	WHR,adjBMI	87
rs2943650	*IRS1*	2	226.8	FAT%,	108
rs4684854	*PPARG*	3	12.5	WHR,adjBMI in women	117
rs6784615	*NISCH-STAB1*	3	52.5	WHR,adjBMI	87
rs6795735	*ADAMTS9*	3	64.7	WHR,adjBMI	87
rs13078807	*CADM2*	3	86.0	BMI	86
rs7638110	*MRPS22*	3	138.9	adolescent obesity (BMI)	118
rs9816226	*ETV5*	3	187.3	BMI	86
rs10938397	*GNPDA2*	4	44.9	BMI	86
rs13107325	*SLC39A8*	4	103.4	BMI	86
rs4864201	*BC041448*	4	131.0	childhood obesity	113
rs11743303	*MAP3K1*	5	55.9	WC in	117

SNP	Locus	Chr	Position (Mb)	Phenotype	Reference
				women	
rs4704220	*COL4A3BP-HMGR*	5	74.8	BMI	[116]
rs2112347	*FLJ35779*	5	75.1	BMI	[86]
rs261967	*PCSK1*	5	95.9	BMI	[112]
rs10478424	*HSD17B4*	5	118.8	WHR,adjBMI in women	[117]
rs4836133	*ZNF608*	5	124.4	BMI	[86]
rs6861681	*CPEB4*	5	173.3	WHR,adjBMI	[87]
rs1294421	*LY86*	6	6.7	WHR,adjBMI	[87]
rs2076592	*BTNL2*	6	32.5	WHR,adjBMI	[87]
rs206936	*NUDT3*	6	34.4	BMI	[86]
rs6905288	*VEGFA*	6	43.9	WHR,adjBMI	[87]
rs2272903, rs987237, rs734597	*TFAP2B*	6	50.9	BMI, WC	[86, 109, 116, 119]
rs9491696	*RSPO3*	6	127.5	WHR,adjBMI	[87]
rs9356744, rs2206734	*CDKAL1*	6	20.8	BMI	[111, 112]
rs1055144	*NFE2L3*	7	25.8	WHR,adjBMI	[87]
rs4240624	*PPP1R3B*	8	9.2	CT hepatic steatosis	[86]
rs17150703, rs7826222	*TNKS/MSRA*	8	9.8-9.9	Extreme obesity, WC	[109, 120]
rs4735692	*HNF4G*	8	76.8	class 1 obesity, overweight	[114]
rs10968576	*LRRN6C*	9	28.4	BMI	[86]
rs16933812	*PAX5*	9	37.0	adolescent obesity (Total Fat Mass)	[118]
rs11142387	*KLF9*	9	72.2	BMI	[111]
rs1211166	*NTRK2*	9	86.5	BMI	[116]
rs7081678	*ZEB1*	10	32.0	WHR,adjBMI	[87]
rs2116830	*KCNMA1*	10	78.3	Extreme obesity	[121]
rs4929949	*RPL27A*	11	8.6	BMI	[86]
rs988712, rs10767664, rs6265, rs2030323	*BDNF*	11	27.5-27.7	BMI, Extreme obesity	[86, 112, 116, 121]
rs3817334	*MTCH2*	11	47.6	BMI	[86]
rs718314	*ITPR2-SSPN*	12	26.3	WHR,adjBMI	[87]
rs7138803, rs7132908	*FAIM2*	12	4.9	BMI, childhood obesity , Extreme obesity	[86, 113, 119]
rs1443512	*HOXC13*	12	52.6	WHR,adjBMI	[87]
rs4771122	*MTIF3*	13	26.9	BMI	[86]

SNP	Locus	Chr	Position (Mb)	Phenotype	Reference
rs9568856	*OLFM4*	13	53.0	childhood obesity	113
rs7989336	*HS6ST3*	13	95.8	class 2 obesity	114
rs534870	*SPRY2*	13	79.9	FAT%	108
rs11847697	*PRKD1*	14	29.6	BMI	86
rs11624704, rs10150332, rs10146997	*NRXN3*	14	77.9-79.0	WHR,adjBMI, BMI, WC	86, 110, 122
rs997295, rs2241423, rs4776970	*MAP2K5*	15	65.8-65.9	BMI	86, 112, 116
rs2531995	*ADCY9*	16	4.0	class 1 obesity	114
rs12444979	*GPRC5B*	16	19.8	BMI	86
rs12597579	*GP2*	16	20.2	BMI	112
rs4788099, rs7359397	*SH2B1*	16	28.8	BMI	86, 116
rs142108, rs1421085, rs1558902, rs8050136, rs3751812, rs9936385, rs9939609, rs9941349, rs17817449, rs12149832	*FTO*	16	52.4	BMI, Extreme obesity, FAT%, childhood obesity	86, 104, 108, 120, 122-124 111-113, 119, {Guo, 2013 #723
rs7503807	*RPTOR*	17	76.2	overweight	114
rs9299	*HOXB5*	17	44.0	childhood obesity	105, 113
rs17700144, rs663129, rs571312, rs17782313, rs12970134, rs2229616, rs2331841, rs6567160, rs12970134	*MC4R*	18	56.0-56.2	BMI, Extreme obesity, childhood obesity, WC in men	86, 111-113, 116, 120, 123, 125
rs2228603	*NCAN*	19	19.2	CT hepatic steatosis	126
rs29941	*KCTD15*	19	39.0	BMI	86
rs2075650	*TOMM40-APOE-APOC1*	19	50.1	BMI	116
rs11671664, rs2287019	*GIPR-QPCTL*	19	50.9	BMI	86, 111, 112
rs3810291	*TMEM160*	19	52.3	BMI	86
rs13041126	*MRPS33P4*	20	50.5	class 1 obesity	114

SNP	Locus	Chr	Position (Mb)	Phenotype	Reference
rs4823006	*ZNRF3-KREMEN1*	22	27.8	WHR,adjBMI	[87]
rs5996074	*SREBF2*	22	40.6	BMI	[116]
rs738409	*PNPLA3*	22	42.7	CT hepatic steatosis, NAFLD	[126, 127]

Chr: Chromosome

1.4.5 Insights gained from GWAS

1.4.5.1 Candidate Genes and non-coding sequences

Most of the GWAS signals found for T2D and obesity map to non-coding sequence.[128] With the exception of a few loci, it is often unclear which gene in a region of association is biologically responsible for the disease risk. Finding candidate genes and causal variants is complicated by the fact that neighbouring genetic variants, even at long ranges of up to ±500 kb, are correlated (or in linkage disequilibrium, LD). A lack of biologically plausible candidate genes can also reflect inadequate understanding of regulatory elements controlling the expression of each gene[129-131], or the lack understanding of the processes involved in the regulation of metabolic homeostasis. Many efforts using fine-mapping and functional approaches are ongoing to characterize pathophysiological mechanisms. In addition, more general studies that characterise regulatory features in the genome, such as the ENCODE study can improve the understanding of genetic regulation at such loci.[131, 132]

For a few loci, the GWAS signal has been directly informative in finding a strong biological candidate gene. The largest signal for T2D in Europeans at the *TCF7L2* (transcription factor 7-like 2) locus, previously found through candidate gene studies[101], has been shown to regulate late events in insulin secretion from pancreatic islet beta cells through animal studies[133] but has also been suggested to act via pathways in the liver.[134] Similarly, alterations in transcript coding sequence of

the ZnT-8 zinc transporter (*SLC30A8*) have been shown to be responsible for the association signal.[135-137] GWAS associations in the glucokinase regulatory protein (*GCKR*), which do occur in the coding sequence, have been suggested to exert effects through glucokinase activity.[137] Another example is Kruppel-like factor 14 (*KLF14*), where variants identified by GWAS cause insulin resistance independently of obesity by disruption of adipocyte function.[138] Additional signals with strong biological candidates include melatonin receptor 1B (*MTNR1B*) and insulin receptor substrate-1 (*IRS1*).[139-141] Similar approaches applied to GWAS for overall obesity and fat distribution have suggested distinct biological processes involved in each. Associations with BMI implicate transcripts involved in hypothalamic function and WHR transcripts influencing adipocyte development and function.[86, 87] These strong biological candidates include *MC4R* and *POMC* for BMI[86, 105] and *TBX* for WHR.[87]

1.4.5.2 Disease Pathways and Relationships between traits

Despite the lack of direct proof of causal variants and transcript at most GWAS loci, their discovery presents a powerful tool to explore aetiological interactions.[85, 90, 142, 143] The relationship between T2D and continuous traits including fasting glucose, insulin secretion, BMI and birth weight has shown that variants altering T2D risk are only partially overlapping with those influencing fasting glucose in healthy non-diabetic individuals.[85, 90] Thus, the processes involved in the physiological regulation of glucose levels are hypothesised to not be identical to those responsible for the development of T2D. This is consistent with the observation that not all individuals with raised fasting glucose levels are at equal genetic risk of developing T2D. By exploring the effects of T2D risk alleles on indices of insulin secretion (HOMA-B) and action (HOMA-IR) in large numbers of non-diabetic individuals[90], it was found that most of the known T2D-risk loci act via effects on insulin secretion, which highlighted the involvement of pancreatic beta cells in T2D genetics.[130]

1.4.5.3 Missing Heritability

The small effect sizes of the risk variants discovered by GWAS mean that in combination the proportion of overall trait variance and proportion of the observed familial aggregation explained is modest. The ~80 T2D variants combined explain ~6% of the variance in disease susceptibility and thus 10-20% of the heritability.[89] Similarly, the variance in BMI attributable to the known common variant signals (~2%)[86] is very small compared to the overall genetic variance estimated previously. The largest signal for T2D, at the *TCF7L2* (transcription factor 7-like 2) locus, is associated with a difference in T2D risk of approximately two-fold between the two homozygote classes[101]. For BMI, the *FTO* (fat mass and obesity associated) locus has the largest effect size (increasing BMI by 0.39 kg/m^2 per risk allele) and an explained <1% of phenotypic variance.[86, 144] Small effect sizes also mean that the genetic variants cannot be considered clinically useful for disease diagnosis.[145]

The unexplained variance and "missing heritability" has sparked studies looking at genetic variants beyond the common SNPs to other types of variants such as low-frequency rare alleles and copy number variants (CNVs). These efforts are undertaken by next-generation sequencing of either exons or whole genomes to extend the surveys of DNA variation. For example, first efforts have suggested that CNVs at neuronal growth regulator 1 (*NEGR1*)[106] and 16p11.2[146] are associated with BMI, but the finding do not add substantially to the overall variance explained. Similarly, a CNV near *KCNIP1*[147] for insulin secretion and 287 CNVs for WHR have been implicated.[148] Unexplained familial aggregation has also been suggested to have arisen from rare alleles that are of relatively recent origin.[149]

Additionally, polygenic risk-score analyses that extend the use of genome-wide associations beyond the genome-wide significance level suggest that accumulated effects explain as much as 50% of the heritability for T2D[150] and ~20% for BMI[151], which is a large proportion of the genetic component. An alternative explanation are

gene-gene or gene-environment interaction effects.[152] The reason why genetic effects, and especially those through common variation, cannot be expected to fully account for familial aggregation, or heritability, can also be attributed indistinct boundaries between genetic and epigenetic effects, as described in the next section.

1.5 Epigenetics of T2D and obesity

The term epigenetics, meaning “above genetics” relates to the study of mitotically stable changes in gene expression other than by changes of the DNA sequence.[153] In this section I will illustrate several lines of evidence that point to epigenetics playing a role in the development of T2D and obesity.

1.5.1 Definition and Types of Epigenetic Marks

The main epigenetic marks are histone modifications (**Figure 1.2A**)[154] and DNA 5-methylcytosine within CpG dinucleotide sites (**Figure 1.2B**), which are also important in genetic imprinting. These marks are directly or indirectly involved in the formation of euchromatin and heterochromatin, which mediates transcriptional activity.[155, 156] This is achieved by condensing the three-dimensional “pearl necklace” structure DNA wrapped around the nucleosome octamers, which are made up of two copies of Histone 2A, 2B, 3 and 4 each (**Figure 1.2A**).[156] Many histone modifications such as methylation, acetylation, phosphorylation, ubiquitination and sumoylation are known, but the most prominent one is methylation on lysines located on the 4, 9 and 7 positions on histone 3. H3K4me3 and H3K9me in a promoter are activating, whereas H3K27me3 and H3K9me deactivate transcription. Acetylation on H3 and H4 is generally associated with transcriptionally active genes[156] These marks have been also shown to act in concert[155], giving rise to ambiguous states that can be changed into the necessary state during development (epigenetic plasticity). Many enzymes have been found that can read, write and erase histone modifications[157], such as histone acyl transferases (HAT) and histone deacetylases (HDAC).

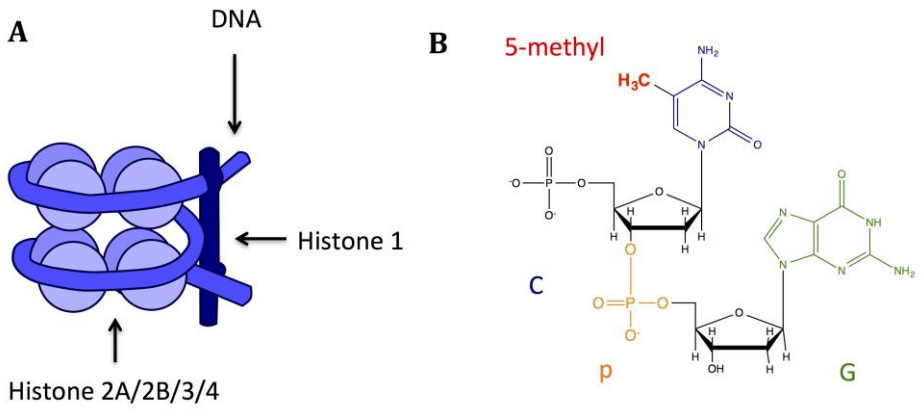


Figure 1.2: The two main components of the epigenetic code. **A:** Two turns of double-stranded DNA (grey) are wrapped around an octamers consisting of two units of histone 2A/2B/3 and 4 each. Histone 1 (what colour) acts as a spacer between these units. **B:** CpG site is a 5' cytosine (blue) followed by a guanosine (green) in the DNA , with the "p" representing the linking phosphate group (orange) between consecutive. The 5-position on the cytosine base can be either methylated (red) or unmethylated.

A third epigenetic pathway is through short RNAs (20-30 bp), such as small-interfering (siRNA) and micro-RNA (miRNA).[158] In contrast to DNA methylation, these are often regarded as a fine-tuning device to control levels of transcription, rather than on/off switches.[158] These can act either by a one-to-one inactivation of a transcript or through catalytic enzymatic degradation ("dicing") through the RNA interference (RNAi) pathway. Moreover, it has been shown that short RNAs can direct DNA methylation and other epigenetic changes.[158]

1.5.1.1 Correlation with transcription

DNA methylation, when occurring in promoter regions and CpG islands, is classically associated with negative regulation of transcription.[159] This has been shown to occur both due to methyl binding domain proteins[160], but also direct weakening of the binding of the transcription machinery.[161] However, this does not necessarily hold for the rest of the CpG sites. Intragenic CpG sites, as well as repetitive elements (short and long interspersed nuclear elements)[162], are found to be mostly methylated[163, 164], and indeed higher levels of intragenic DNA methylation are correlated with higher levels of gene expression.[165]

1.5.1.2 Mitotic stability and propagation

While epigenetic changes can, at least in part, remain stable through mitotic cell divisions[166], their meiotic heritability across multiple generations is not well established. Hence, the definition “mitotically stable differences” instead of “heritable differences” is often used.[167] In humans, meiotic epigenetic heritability is complicated by the fact that twice during development (during gametogenesis and shortly after fertilization of the egg cell), there is a large genome-wide erasure of epigenetic marks, which are then later re-established in somatic cells.[168] However, this erasure can be incomplete such as in genetically imprinted loci, where methylation that distinguishes maternal from paternal chromosomes persists into mature cells. Genetic imprinting not only gives rise to this parent of origin effect, but also ensures that transposons are silenced during gametogenesis.[168] Examples of human imprinted genes include *H19* (maternal allele inherited)[169] and *IGF-2* (paternal allele inherited).[170] Random or incomplete erasure can result in imprinting failures that give rise to severe diseases[171].

1.5.1.3 Transgenerational inheritance

While the evidence for transgenerational epigenetic inheritance is still low and controversial in humans[172, 173], it has been shown to exist in other species. [174-176] When an epigenetic mark is inherited, it is known as an epiallele, and if it is influenced by the environment or erased after a few generations, it can be referred to as a metastable epiallele. One example relevant to metabolism is the agouti viable yellow (A^{vy}) mouse metastable epiallele[174]. This transgenic species can take on a different coat colour (from yellow to brown) depending on maternal nutrition and environmental exposures, which are then also transmitted through several generations[177]. The A^{vy} allele is amongst the rare cases of a mammalian epiallele, which include those in *Axin* and *Cabp*, but all these are still caused by an artificially inserted retrotransposon rather than naturally occurring. [177] Moreover, known

heritable epialleles that interact with the environment are known in plants such as maize inbreeds[175] and *Arabidopsis.*[176]

1.5.1.4 Inheritance via meQTL, CpG-SNPs and haplotype-specific methylation

Apart from true epialleles, a possible mechanism for meiotic inheritance of epigenetic patterns is made possible by using genetic variants as a vehicle. Methylation quantitative trait loci (meQTLs) have been found to be abundant across the genome and many cell types.[178, 179] These may either alter DNA methylation at nearby sites through a *cis* effect, for example by altering binding of the methylation machinery, or by directly changing a CpG site.[180] In the latter case, multiple CpG-altering SNPs acting in the same direction might give rise to haplotype-specific methylation (HSM). In that situation, the methylation status of a region is directly determined by the status of allelic variants that create or abrogate methylation sites.[181]

1.5.2 Epigenetic Studies of T2D and Obesity

1.5.2.1 Animal Studies

Epigenetic processes represent strong candidates for mediating the effects of T2D and obesity. Animal models have provided a growing body of data and experimental support for a causal link between early nutritional status and epigenetic changes that predispose to T2D and obesity risk in later life.[182] Despite identical genetic make-up, the fur coat colours of agouti viable yellow (A^{vy}/a) mice range from yellow to brown, and those colours are correlated with adult body weight.[174] The variation in phenotypes is caused by an epigenetic modification of the A^{vy} allele passed through the female germ line that results in stable intergenerational transmission.[174] This causes both phenotypes, as the agouti protein controls both fur colour and eating behaviour when overexpressed.[174]

Nutrition can also affect methylation status of DNA in rats. A 70% reduction in total nutrient intake during pregnancy led to a phenotype similar to human metabolic

syndrome due to hypermethylation and decreased expression of the glucocorticoid receptor (*GR)* and peroxisome proliferator activated receptor alpha (*PPAR*A).[183] This effect has been shown to be heritable to two generations, but did not stay stable to the third generation.[184] Hypermethylation of the *POMC* promoter can be induced by overfeeding newborn rats.[183] This gene is normally induced by leptin and insulin, but methylation prevents this upregulation, which then results in increased incidences of obesity in later stages of life. An association between hypermethylation of an Alu-repeat element was also found to associate with childhood obesity in humans.[185] Reduced methylation of the *MC4R* promoter, another gene important in body weight regulation, due to a high fat diet was also observed in mice.[186] A search for obesity markers in obese human individuals using a hypocaloric diet intervention revealed significant methylation differences in the *ATP10A*, *CD44* and *WT1* loci of peripheral blood mononuclear cells between high- and low responders.[187]

1.5.2.2 Human Population Studies

The foetal origins hypothesis established the notion of "metabolic programming" whereby nutritional and other exposures during early life generate long-term changes that later predispose to T2D and cardiovascular disease.[52] This hypothesis is supported by epidemiological data linking early life events to disease risk in late life. An example for human epigenetic inheritance is the Dutch Hunger Winter Families study. Children who had experienced intrauterine famine showed less blood DNA methylation in the *IGF2* locus and high tendency to develop obesity later in life that was also inherited to the third generation. In the ALSPAC cohort it was found that in pre-adolescents, paternal smoking associates with high BMI in sons, but not daughters (see review by Morgan et al.[168]).

Further, an early study in Pima Native Americans in Arizona (a population with high prevalence of T2D and obesity) showed that maternal diabetes during pregnancy associated with obesity and diabetes in the offspring.[50] The offspring of mothers who

have T2D during pregnancy are at substantially higher risk of developing both T2D (45% vs. 1.4%) and obesity (58% vs. 17%) than those born to women who are non-diabetic during pregnancy.[50, 51] This is thought to be caused by higher glucose levels in utero, which also leads to higher birth weight.[188] A discordant sibling study showed that subjects who had experienced intrauterine maternal diabetes have a much higher risk of obesity and diabetes compared to their siblings born before maternal diabetes. This shows that the effects are present even under similar genetic backgrounds[189] and are unlikely to completely reflect genetic transmission, since the distinction is preserved in children born to the same mother.[189] Thus, the intrauterine environment is an important determinant of T2D and obesity risk later in life. The increased risk of diabetes in female offspring of diabetic mothers sets up the potential for an amplification of diabetes prevalence over successive generations. This is consistent with reports of greater transmission of T2D and obesity by mothers than fathers.[51, 52] Epigenetic changes, such as DNA methylation, are a good candidate for the mechanism behind the foetal origins theory.

1.5.2.3 Human Candidate Epigenetic Studies

There have been few studies exploring the epigenetic component in the risk of T2D or obesity (**Table 1.4**). They have mostly focused on characterizing the methylation status of selected CpG sites in candidate genes, though the first “epigenome”-wide studies have now been performed. Similar to genetic candidate gene studies, these have often been poorly powered and lacking replication. Main challenges in the research are the high cost and complexity of epigenetic analysis compared to genetic ones, and by tissue differences in DNA methylation. This means findings obtained from accessible samples (e.g. peripheral blood) may not be representative of those present in more disease-relevant tissue, or secondary effects consequential of the phenotype of interest. The findings thus need to be replicated in the largely inaccessible, relevant tissues, such as the pancreatic islet or hypothalamus.

An example of the challenges involved in candidate gene methylation studies comes from analyses targeting the leptin-responsive gene *POMC.*[185] *POMC* encodes several key neuropeptides and hormones[70] and common as well as rare mutations in *POMC* are linked to obesity.[190] A hypermethylated CpG site at the intron2/exon3 boundary of *POMC* identified in human peripheral blood leucocytes (PBL) has shown replicated association with childhood obesity.[185] This hypermethylated position overlaps with a putative binding site for the histone acetyltransferase P300 complex, which is known to be involved in chromatin acetylation and gene activation, which potentially affects *POMC* expression.[185] The suggestion is that similar methylation effects are present within the POMC neurons, for which methylation in PBLs is a proxy.

1.5.2.4 Epigenome-wide association studies

The current status of much epigenetic research in humans is similar to that of genetics at the beginning of the GWAS era. Previously, underpowered and unreplicated candidate have led to largely inconclusive results, and there is no estimate of the extent to which risk of T2D and obesity is reflected by epigenetic variation. The field is now carrying out large-scale, global studies of epigenetic marks to get a systematic view of the epigenetic contribution to disease pathogenesis. These types of studies have been termed epigenome-wide association studies (or EWAS), following the term for genetic studies. While some of the smaller studies are targeting disease-relevant tissues, others focus on whole blood for practical and economical reasons.

These studies have become feasible through the development of epigenome-wide DNA methylation profiling platforms such as the Illumina HumanMethylation450 BeadChip (450k array)[191], which take into account improved knowledge of the variability of the genome. These array-based assays are increasingly giving way to a

range of methods that couple bisulphite treatment with next-generation sequencing to capture regions of differential methylation in a more unbiased fashion.

Despite these technological advances, designing an appropriate EWAS remains challenging, as these studies combine negative aspects from both epidemiological studies and genome-wide testing. Epigenetic marks can change over time, which means that environmental influences and reverse causation due to the disease of interest or its treatment[192] can lead to confounding. Moreover, differences in tissues patients, sample collection protocols and laboratory procedures are all likely affect results and make comparison across studies difficult. Collecting sufficient numbers of samples from disease-relevant tissues and/or cell types[192] is another challenge.

These are potential reasons for why there are only few EWAS for T2D and obesity that have been carried out on relatively modest scale (**Table 1.4**). The largest study before the development of the 450k array employed a pooled design to interrogate genome-wide methylation in whole blood from a sample of ~1150 T2D cases and controls, which revealed an enrichment for genetic loci previously implicated in T2D risk.[193] Of particular interest in this study were CpG sites in the first intron of the *FTO* gene, where methylation levels correlated with genotype at nearby BMI-associated SNPs in blood.[181] Associations of the *FTO* risk allele with hypermethylation have previously discovered and replicated in several studies[181, 193, 194], but whether methylation represents the causal link between the *FTO* risk allele and T2D via obesity is still debatable.[193]

Table 1.4: Methylation candidate studies and EWAS

T2D					
Candidate genes (HGNC Symbol) / EWAS (a)	**Phenotypes**	**Case control / quantitative trait**	**Sampled tissue**	**Replication**	**Reference**
CCL2	T2D	c/c (b) 32/15	peripheral blood mononuclear cells		[195]
INS	T2D	c/c: 9/48	pancreatic islets		[196]
PDX1	T2D	c/c: 9/55	pancreatic islets		[197]
PPARGC1A	T2D	c/c: 10/9	pancreatic islets		[198]
EWAS	T2D	c/c: 17/17	skeletal muscle		[199]
EWAS	T2D	c/c: 40/40	whole blood		[181]
EWAS	T2D	c/c: 710*/459*	whole blood	c/c: 198*/233*	[193]
EWAS	T2D	c/c: 11/5	pancreatic islets		[200]
T2D					
EWAS	T2D	34/15	pancreatic islets	34/15	[201]
EWAS	Fasting Glucose, Fasting Insulin and HOMA-IR	544	CD4T	293	[202]
GIPR	T2D	93/93	whole blood	93/93	[203]
IRS-1	T2D	48/48	can only get abstract	48/48	[204]

Obesity					
Candidate genes (HGNC Symbol) / EWAS (a)	**Phenotypes**	**Case control / quantitative trait**	**Tissue**	**Replication**	**Citation**
ALOX12, ALPL, BCL2A1, CASP10, CAV1, CCL3, CD9, CDKN1C, DSC2, EPHA1, EVI2A, HLA, IRF5, KRT1, LCN2, MLLT4, MMP9, MPL, NID1, NKX31, PMP22, S100A12, TAL1, VIM	BMI, fat mass, lean mass, height	qt (c): 178	cord blood		[205]
KCNQ1OT1, H19, IGF2, GRB10, MEST, SNRPN, GNAS	BMI discordance in twins	c/c: 16/16	saliva		[206]
MCHR1	BMI	qt: 49	whole blood		[207]
POMC	obesity	c/c: 71/36	whole blood	c/c: 54/100	[185]
IL8, NOS3, PIK3CD, RXRA, SOD1	fat mass, %fat mass	qt: 78	umbilical cord tissue	qt: 239	[208]
SLC6A4	BMI, weight, waist circumference	qt: 168	peripheral blood leukocytes		[209]
TACSTD2	fat mass	qt: 94	whole blood	qt: 161	[210]
EWAS	BMI	qt: 64	lymphocytes		[211]
EWAS	obesity	c/c: 7/7	peripheral blood leukocytes	c/c: 46/46	[212]
EWAS	obesity	c/c: 23/24	whole blood		[194]
EWAS	BMI	479	whole blood	339;1789	[213]

a: epigenome-wide association study b: case/control c: quantitative trait *: pooled samples

These early studies suggest that the methylation status may be relevant to the T2D phenotype, and that epigenetic effects at *FTO* may, at least in part, be driven by underlying variation in the DNA sequence. If so, *FTO* would represent an example of haplotype-specific methylation (see Chapter 1.5.1.4, p. 24). From a genetic (or GWAS) perspective, the epigenetic association would then be a subset of genetic association signals at which the downstream effects are mediated by DNA sequence dependent changes in local methylation.

1.5.3 Publically available epigenome projects

Whilst there is evidence that epigenetic effects are likely to play a role in the pathogenesis of T2D and obesity, and contribute to the observed familial aggregation, experimental studies that conclusively define the key loci responsible in disease-relevant tissues have yet to be performed (**Table 1.4**). It is also worth noting that existing studies have focused almost exclusively on DNA methylation, leaving other components of the epigenetic machinery (such as chromatin state) largely unexplored. Given limited access to disease-relevant human tissues, progress will depend on efforts such as the "International Human Epigenome Consortium (IHEC)", which is generating reference epigenomes (including detailed methylome maps) for a panel of human tissues and cell types.[192] In 2012, the ENCODE project published the largest catalogue of epigenomic maps to date, combining a vast quantity of information comprising accessible chromatin (125 cell types), profiles of 12 different histone modifications (46 cell types) and 119 transcription factors (72 cell types), DNA methylation information (82 cell types), RNA expression levels (15 cell lines), and chromatin interaction maps (4 cell lines)[131], providing important insight into the complexities of gene regulatory networks.[132] These, and other studies, can empower design future methylation arrays and EWAS by providing tissue-specific maps of methylation correlation structures and variably methylated sites.

1.6 Thesis Aims

This introduction has highlighted evidence to support the notion that individual predisposition to T2D and obesity reflects a complex mix of genetic, epigenetic and environmental influences. Despite recent progress in the field of epigenomics, the mechanisms driving these interactions, such as DNA methylation, remain poorly understood. Thus my DPhil project focuses on large-scale assessments of the roles of DNA methylation in obesity and T2D to characterise the epigenetic component. Specific aims are:

1) To investigate the effect of genetic variants on DNA methylation in tissues relevant to T2D and obesity, such as adipose tissue (Chapter 3).
2) To identify important study design aspects in an EWAS pilot study (Chapter 4).
3) To optimise the analysis of epigenome-wide methylation data, including quality control, normalisation and correction for confounding factors (Chapter 4).
4) To apply the findings from Aim 1 and Aim 2 to discover associations of methylation with T2D and obesity through methylation profiling in blood from a large population-based cohort (Chapter 5).
5) To investigate enrichment of epigenetic effects at previously identified genetic regions associated with obesity and T2D traits (Chapter 6).

Chapter 2 Methods

2.1 The EpiMigrant Study

The EpiMigrant study is a large-scale study of DNA methylation in migrant South Asians. The work described in Chapter 4 and Chapter 5 was based on three DNA methylation data sets within the study.

2.1.1 Samples and Phenotypes

All samples were selected from the selected from the London Life Sciences Population (LOLIPOP) study, which has previously been ethically approved by the West London Ethics Committee. Participant characteristics are shown in **Table 2.3**. Three data sets were used:

1. A longitudinal pilot study of 90 samples (15 controls, 15 T2D cases on medication (T-T2D), 15 T2D cases not on medication (U-T2D), all measured at baseline and 7.0±0.9 years)
2. A prospective case-control study of 2,680 samples (1,080 T2D cases, 1,600 controls) with all samples sampled before T2D onset
3. A replication dataset of 36 samples measured in duplicate.

Table 2.1: Participant characteristics for the EpiMigrant Study

Participant characteristics	
N	2680
Country	UK
Ethnicity	South Asian
Study Design	Prospective T2D Case/Control
Age (yrs)	51.0 (10.1)
Sex (M)	67.7%
Fasting glucose (mmol/L)	5.2 (0.6)
HbA1c (%)	5.5 (0.5)
Fasting insulin (IU/L)	12.6 (10.2)
% Fasting	100%
Weight (kg)	76.2 (13.8)
Height (cm)	166.1 (9.2)
Body mass index (kg/m^2)	27.6 (4.4)
Waist circumference (cm)	97.3 (11.2)
Waist-hip ratio	0.95 (0.07)
Alcohol (g/d)	5.6 (12.5)
LDL cholesterol (mmol/L)	3.4 (0.9)
HDL cholesterol (mmol/L)	1.3 (0.3)
Total cholesterol (mmol/L)	5.4 (1.0)
Triglycerides (mmol/L)	1.7 (1.1)
C-reactive protein (mg/L)	4.2 (7.2)
Systolic BP (mmHg)	131.6 (18.9)
Diastolic BP (mmHg)	81.8 (10.7)
HT**	33.5%
Treated HT	24.40%
CHD***	7.8%
T2D****	0.0%
Physically Active	28.7%
Smoking	
Never smoked	82.7%
Ex-smoker	8.5%
Current smoker	8.8%

Values indicate mean (SD) for each trait. *both fasting/non-fasting subjects, **HT (Hypertension): SBP ≥ 140mmHg, or DBP ≥ 90mmHg, or who were taking anti-hypertensive or blood pressure-lowering, ****T2D (Type 2 Diabetes): physician diagnosis or HbA1c>6.5%

2.1.2 CpG Methylation Data

Genomic DNA was extracted from peripheral blood and analysed in batches of 288 samples. Within each dataset, all samples were randomised and initial and repeat sample analyses carried out in separate batches. Methylation was quantified with the Illumina Infinium HumanMethylation450 (450k) array using the published protocol [191] with the following changes: 1ug of DNA was used as starting material and an elution volume of 14 ul after bisulphite conversion using the EZ-96 methylation kit (Zymo). The microarray chips were imaged using an Illumina HiScan scanner. Methylation was quantified using the beta value, which is defined by the following intensity (*I*) ratio for each probed locus:

$$beta = \frac{I(methylated)}{I(methylated)+I(unmenthylated)+100} \quad (1)$$

2.1.3 Outliers and Outlier Rate

To assess the quality of the measurements at a marker, the outlier rate was calculated. For each methylation marker, outliers measurements are detected based on the interquartile range (IQR), such that beta values are considered as outliers if they fall below quartile 1 – 1.5 × IQR or above quartile 3 + 1.5 × IQR. Outlier rates are calculated as the number of outlying beta values divided by the number of non-missing beta values. The outlier rate was used to test detection p-value thresholds and for marker exclusion.

2.1.4 Data Normalisation

All normalisation methods investigated are summarised in **Table 2.4**. Data normalisation was carried out to decrease experimental variability.

Table 2.2: Normalisation methods for the 450k array

Name	Aim	Input	Categories	Method
QN (I)	Normalise between arrays	Intensities	6 (Colour, Type)	Quantile Normalisation
QN (B)	Normalise between arrays	Betas	3 (Type)	Quantile Normalisation
Illumina	Normalise between arrays	Intensities, Control Probes	NA	Normalisation to Control Probes
SWAN	Normalise within arrays	Betas	3 (1-3 CpGs)	Quantile Normalisation
PBC	Normalise within arrays	Betas	2 (Type)	Rescaling to Kernel Density Estimate
BMIQ	Normalise within arrays	Betas	2(Type)	Transformation to methylation probability, fitting of quantiles
Tost	Normalise within arrays	Betas	2 (Type)	Quantile Normalisation of Type II based on Type I

Two quantile normalisation strategies were developed to normalise samples across arrays:

1. Quantile normalisation of beta values (QN[B]). For each sample beta values are ranked by increasing beta; the value of each observation is then assigned the mean beta (across samples) for that rank using the R function *normalizeQuantiles* (limma).[214]
2. Quantile normalisation on intensity values (QN[I]). The intensities are split into subsets based on probe type (Type I and Type II) and colour channel (green and red). This results in 6 categories (Type I Methylated Green, Type I Unmethylated Green, Type I Methylated Red, Type I Unmethylated Red, Type II Unmethylated, Type II Methylated) which are separately ranked by increasing intensity and then assigned the mean intensity for each rank using the R function *normalizeQuantiles* (limma).[214]

In addition, five other published methods were evaluated, which provide either between-array or within-array normalisation:

3. Illumina Control Probe normalisation (Illumina). Control probe pairs, which have no CpG sites in the target DNA sequence, are used for normalisation. Probe intensity in the given sample is multiplied by a constant normalization factor and divided by the average of normalisation controls in the probe channel. The method is implemented in the minfi package.[215]
4. Subset within-array normalisation (SWAN). First, an average quantile distribution is created by selecting N Type I and II probes that have 1, 2 and 3 underlying CpGs. The value of each of the 3N pairs of observations is subsequently normalised using the quantile procedure.[216]
5. Peak-based correction (PBC). Type I and II probes are rescaled separately to the estimated peaks in the kernel density distribution of the M values. Subsequently, both distributions are rescaled to the initial Type I probe distribution and M values converted to beta values.[217]
6. Beta Mixture Quantile normalisation (BMIQ). Similarly to PBC, a beta mixture-model is applied to assign probabilities for methylation, followed by two transformations (transformation of probabilities into quantiles and methylation-dependent dilation).[218]
7. Subset quantile normalization (Tost Pipeline): Beta values for Type I and II are separated and then only Type I probes are filtered for detection p-value > 0.01. From the Type I probes, reference quantiles are computed, which are then used to estimate reference quantiles for the appropriate number of Type II probes. Type I and II beta values are then normalised to the respective reference quantiles.[219]

2.1.5 Marker Quality Control

Many markers on the 450k array are measured with probes that are potentially confounded by technical variability. The following categories were investigated:

2.1.5.1 Cross-Hybridising Probes

Previously, several methylation probe sequences that map to more than one genomic location (N=39,963) were identified by Price et al.[220] To reassess the extent the further extent of cross-hybridising on the Y chromosome, probe sequences from the Illumina annotation were converted into the BED format and mapped to the bisulphite-converted genome hg19 (4 strands, alternative haplotypes removed) using blat with the following flags: -oneOff=1 -minMatch=1 -stepSize=5 -noHead -tileSize=11. Results were filtered for >=45 matching bases, no gaps, 90% sequence identity and the 50th base matching.

2.1.5.2 Probes with Single Nucleotide Polymorphisms (SNPs)

Many probes on the 450K array harbour Single Nucleotide Polymorphisms (SNPs). These were flagged if at least one SNP was located within the 50 base pair probe-sequence and had a MAF above 1% based on genotyping data of 1,909 Indian Asians (71% of individuals in this study).

2.1.5.3 Probes with a low Call-Rate

Call rates based on detection p-values were determined for each methylation marker. Any markers with a call rate below 98% were flagged as low call rate markers.

2.1.5.4 CpA Probes

The 450k array contains a number of probes that measure methylation at CpA rather than CpG sites (N=2,995) sites, which are flagged in subsequent analyses.

2.1.6 Control Probe Adjustment

Intensity values from the 450k array control probes were used to adjust for technical bias. Control probe intensities (excluding negative control probes) were obtained using *minfi*[215] and a principal component analysis was performed. The principal components were then included as linear predictors in regression models to adjust the beta values measured at each CpG site for technical bias.

2.2 Local Correlation Structure

Local correlation was determined for all possible pairs of autosomal markers up to 5,000 bp apart. This distance was chosen as it represents about double the distance of reported correlation of methylation.[163, 179] Distance between markers was based on the annotated position of the CpG sites on the forward strand. Pearson correlation coefficient between marker pairs were calculated based on beta values (raw) and residuals derived from models (7) and (8). A large proportion of methylation markers show very little variation, which limits their ability to yield high correlation coefficients. To reflect the effect of genomic distance on correlation more appropriately the 5% most variable markers (based on raw beta values) were selected. The correlation is represented graphically on a continuous scale using a sliding 300 bp mean. To demonstrate that adjustments preferentially reduced correlation between markers with greater distance the difference in correlation coefficients per base pair distance (between two different adjustments) was calculated.

2.3 Permutation Testing and Global Correlation Structure

To increase the number of permutations that are computationally permissible a two-step approach was employed. First, a linear regression of beta value against the corresponding linear predictors was carried out and the residuals extracted. Second, T2D disease labels were randomly re-assigned followed by a logistic-regression predicting disease-status as a function of the previously derived residuals.

2.4 Spike-in Simulations

As described above, disease labels were randomised to generate 10 permutated EWAS datasets. From each permuted dataset 100 markers are randomly selected and spiked. For each “spike-marker” raw beta values of samples with a case label are increased by a defined proportion of the standard deviation of the respective marker. Based on these spiked beta values intensity values for the methylated and

the unmethylated probe were calculated, such that for half of the spiked probes the methylated intensity is changed and for the other half the unmethylated intensity is changed. Negative intensity values resulting from this process are set to zero. Each permuted dataset is then processed on the basis of the analysis pipeline presented here and the 100 spiked markers are ranked by their association p-values with respect to all other markers in the permuted dataset. All autosomal markers were included in this analysis. The 10 permuted datasets give rise to a total of 1,000 ranks for 1,000 spiked markers.

2.5 Statistical Methods

2.5.1 Regression Models

The *lm()* function in R was used to fit the following linear models to test for association with T2D case/control status (Eq. 2), each of the potential confounders (e.g. age, gender and technical confounders, (Eq. 3), and T2D status adjusted for all confounders (Eq. 4).

$$y = \beta_o + \beta_{T2D} x_{T2D} + \varepsilon \quad (2)$$

$$y = \beta_o + \beta_{conf} x_{conf} + \cdots + \beta_{conf,n} x_{conf,n} + \varepsilon \quad (3)$$

$$y = \beta_o + \beta_{T2D} x_{T2D} + \beta_{conf,1} x_{conf,1} + \cdots + \beta_{conf,n} x_{conf,n} + \varepsilon \quad (4)$$

where y is the methylation score at each probe set and the error term is $\varepsilon \sim N(0, \sigma^2)$ for each marker independently. A t-statistic is then calculated based on the β regression coefficient and the standard error and the corresponding p-value for $H_0: \beta_{predictor} = 0$ is then calculated. Genomic inflation of associations was measured as

$$\lambda = \text{mean}\left(\frac{\beta^2}{\text{var}(\beta)}\right) \quad (5)$$

where β is the regression coefficient.

For paired measurements, such as those over time or between replicates, first the difference in beta was calculated as Δbeta = $beta_2$ – $beta_1$ for each marker. For the assessment of technical confounders, effects of replicated status were compared analogously to Eq. 4 (above) as

$$y = \beta_o + \beta_{rep}x_{rep} + \beta_{conf,1}x_{conf,1} + \cdots + \beta_{conf,n}x_{conf,n} + \varepsilon \quad (6)$$

where y is the difference in beta, Δbeta.

2.5.2 Principal Component Analysis

All principal component analyses (PCA) were performed on the 450k methylation array data to detect biological clustering and outliers for all samples in the study and all markers that had no missing data (>60%). Three PCAs were performed to analyse the 450k methylation array data.

1. PCA of control probe intensities. Control probe intensities were retrieved using minfi and Principal Components (PCs) successively added as predictors until no further statistical inflation was observed in the downstream regression-analyses.
2. PCA of residuals after quantile normalisation and control probe adjustment. Residuals were extracted for all markers after linear regression (7) of the beta value. To identify factors driving global correlation structure the corresponding PCs were linked to multiple phenotypes (age, gender, white-blood cells, etc.) using linear regression. P-values of association (between the PCs and the phenotypes) were Bonferroni-corrected and plotted on the $-\log_{10}$ scale as a heatmap.

$$y = \beta_o + \beta_{PC1}x_{pc1} + \cdots + \beta_{PC30}x_{pc30} + \varepsilon \quad (7)$$

3. PCA of residuals after quantile normalisation and adjustment for technical and biological biases, including age, gender as well as estimated and total white

blood cell counts (WBC). Residuals were extracted for all markers after linear regression (8) of the beta-value. Corresponding PCs were included in the final regression model.

$$y = \beta_o + \beta_{age}x_{age} + \beta_{gender}x_{gender} + \beta_{WBCest}x_{WBCest} + \beta_{WBCtot}x_{WBCtot} + \beta_{age}x_{age} + \beta_{PC1}x_{pc1} + \cdots + \beta_{PC30}x_{pc30} + \varepsilon \quad (8)$$

Chapter 3 DNA Methylation in subcutaneous abdominal adipose tissue

3.1 Introduction

Variants in the primary DNA sequence that associate with methylation at nearby CpG sites offer a potential biological mechanism of action for disease associated SNPs. In this chapter, I investigated epigenome-wide DNA methylation in 38 abdominal SAT samples, which is an accessible, metabolically active fat depot[221], for associations with MetSyn phenotypes and nearby genetic variants.[178, 179]

3.1.1 Relevance of Adipose Tissue for T2D and Obesity

With epigenetic studies, tissue availability is generally a problem and many studies on CpG methylation have been carried out in blood[181, 185, 193] but abdominal SAT is thought to represent a metabolically active tissue that is accessible and of relevance to obesity-related diseases.[221]

Subcutaneous adipose tissue is found in a region below the skin known as the hypodermis. It is a type of white adipose tissue that acts as energy storage as well as a thermal insulator, and is characterised by a single droplet of lipid within the cell.[222] This type of tissue is an active part of the endocrine system, secreting leptin, adiponectin and resistin, and having receptors, amongst others, for insulin.[223, 224] Other than visceral fat, which is located around organs in the abdominal cavity, this type of fat poses a lesser health risk and has even been suggested to be protective of obesity-related diseases.[223]

The study of DNA methylation in SAT is interesting, as epigenetic processes play an important role in adipogensis: in both the establishment of the identity of the white adipose tissue precursors from the mesoderm as well as in terminal adipocyte

differentiation.[221] Understanding factors driving inter-individual variation in this tissue may help understand the differences in disease risk.

3.1.2 Methylation Quantitative Trait Loci (meQTL)

Genetic variants that associate with DNA methylation at CpG sites (methylation quantitative trait loci, meQTLs) offer a potential biological mechanism of action for disease associated SNPs: Disruption of methylation patterns by SNPs can either directly affect transcription at regulatory sites[225], or indirectly through differential splicing at alternative promoters.[226] Moreover, as the methylation marks are generally not inherited across generations in humans[171], meQTL provide a mechanism for transgenerational transmission of methylation patterns.

A number of studies have been carried out investigating the associations between genetic variants, DNA methylation and gene expression. An early study using methylation sensitive SNP analysis for 12 tissue samples, including peripheral blood leukocytes, kidney, brain, placenta and buccal cells, provided evidence for allele-specific DNA methylation (ASM), and demonstrated that most of the associations occurred at short distances in cis. [227] The results were confirmed another study of ASM in blood samples of 5 twin pairs.[228] With the release of the Illumina 27k array, larger studies were undertaken, such as those of 77 HapMap lymphoblastoid cell lines (LCL)[179], as well as two studies of ~150 individuals each in brain tissues.[178, 229]

The field further progressed with the development of the 450k array, leading to studies with higher sample size and first attempts to estimate the heritability of DNA methylation. A study of European and African HapMap lymphoblastoid cell lines showed that there are many population-specific variations in DNA methylation, of which many have an underlying genetic basis.[230] Another family-based study showed that the average heritability of DNA methylation is estimated as 0.187, with strong evidence for *cis*-meQTLs driving the heritability.[231] Lastly, a study was carried out by

the MuTHER consortium of skin, blood and adipose tissue in 648 female twin pairs, allowing for estimation of heritability.[232] The study reproduced the high heritability of methylation (median 0.34) and found that as much as 28% of CpGs were associated with a SNPs.[232] Moreover, the meQTL were enriched for SNPs previously associated with metabolic trait and disease loci:[232] There were 2,768 SNPs within the 74,174 SNPs associated with methylation sites and proxies ($R^2 > 0.8$) that overlapped disease loci in the National Human Genome Research Institute GWAS catalogue (1.7-fold enrichment, $p< 10^{-4}$).[232]

3.1.3 Aims

My work in this chapter followed three main aims:

1) To detect associations of methylation with MetSyn and other phenotypes in adipose tissue
2) To find if meQTLs exist in adipose tissue
3) To investigate if the found meQTLs are enriched for eQTLs or loci previously implicated in disease

3.2 Materials and Methods

3.2.1 The MolOBB Methylation Study

The 38 abdominal subcutaneous adipose tissue (SAT) samples investigated were collected for the MolOBB study, which is part of the MolPAGE project. The case/control study design for MetSyn (for definition, see Chapter 1.2.3 p. 5)[6] involves 20 cases (10 male, 10 female) and 20 controls (10 male, 10 female).

3.2.1.1 Samples and Phenotypes

Abdominal fat biopsies were collected from the Oxford Biobank [233] and analysed for DNA methylation by Epigenomics AG (Berlin, Germany). One male case and one male control were excluded from the analysis due to non-European descent and

missing phenotype data. This left 19 MetSyn cases and 19 control samples for inclusion into the meQTL analysis (**Table 2.1**). The National Research Ethics Service, Oxfordshire REC C (REC reference: 08/H0606/107) approved the study and informed consent in writing was collected from all participants.

Table 3.1: Participant characteristics for the MolOBB study

Participant characteristics	Male cases	Female cases	Male controls	Female controls
Sample size (N)	9	10	9	10
Age (years)	48 (3)	50 (6)	47 (5)	51 (5)
BMI	32.9 (4.9)	31.9 (6.4)	25 (2.6)	28 (5.5)
Waist (cm)	112 (14)	103 (12)	89 (6)	90 (11)
Hip (cm)	113 (10)	113 (12)	100 (4)	106 (10)
HDL (mmol/l)	0.85 (0.13)	1.19 (0.25)	1.36 (0.21)	1.68 (0.18)
TG (mmol/l)	2.6 (1.8)	1.9 (0.3)	0.9 (0.2)	1.2 (0.3)
Glucose (mmol/l)	5.7 (0.5)	5.6 (0.5)	5.2 (0.4)	5 (0.3)

Values indicate mean (SD) for each trait

3.2.1.2 Differential Methylation Hybridisation Data

The DNA from abdominal SAT samples was extracted by Amy Barrett and DNA methylation status was assayed at 51,195 genome-wide regions using Differential Methylation Hybridisation (DMH, Epigenomics AG, Germany). The DMH method is based on a methylation sensitive restriction digest by hybridisation onto a custom microarray as described by Huang et al.[234] Briefly, genomic DNA was digested with methylation unspecific restriction enzymes[235], with short fragments removed (<80 bp) in the purification step. Next, DNA oligomer adapters were ligated to restricted DNA fragments, which were then digested using four different CpG methylation specific restriction enzymes.[235] In a subsequent PCR step, only uncut (unmethylated) fragments were thus amplified, as the digested (methylated) DNA does not have both 5' and 3' adapters for PCR primer binding.[234] PCR products were then hybridised to the custom-designed DMH chip based on the CustomSeq microarray (Affymetrix, USA). A previous study showed both reproducibility and validation of the

methodology.[236] The assay design covers 430 differentially methylated genes, 13,500 CpG-rich promoter regions and 13,650 CpG-rich regions within gene bodies.[235]

The methylation scores for each individual as provided by Epigenomics AG were based on two averaged DMH microarray measurements for each probe set, and all arrays were randomized across three batches. Intensity values of each probe were averaged over the two duplicate chips per sample. The methylation score for each probe set is defined as the difference between the median log2-transformed probe intensities for each probe set and the median log2-transformed chip-internal signal control probe intensities.[235] Epigenomics AG also carried out two additional DMH experiments containing 0% and 100% methylated DNA[235] in order to calibrate the methylation score value to the 0 % and 100% methylated DNA.[237] As the methylation measurements could result in measurements lower than these standards, the observed methylation scores ranged between 1.20 and 2.75. Subsequently the provided matrix of methylation scores was quantile-normalised across individuals and any DMH probe sets that contained SNPs with minor allele frequency (MAF) > 5% in the 1000 Genotypes CEU Interim 2010 data[238] was excluded. The exclusion step was included to reduce allele bias in the hybridisation step and the lowered stringency of MAF > 5% excludes rare variants from the filtering.

3.2.1.3 Genotype Data

302,765 genetic markers from the 38 individuals were genotyped using the Illumina 317k BeadChip platform (Illumina Inc., USA) at the Oxford Genomics Centre, UK. Quality control criteria included (i) Hardy-Weinberg p-value ≤ 5.7x10-7 and (ii) MAF > 1% and genotype rate > 99%, or > 95% if MAF > 5%. Using IMPUTE (v2.1.2)[239] and the 1000 Genomes Interim 2010 haplotypes as a reference panel[238], 11,116,176 SNPs were imputed. Quality control was carried out by applying a filter for effective minor allele count (MAC) > 5% (2 * MAF * IMPUTE Info Score *38 > 5)[86, 87], resulting

in 5,227,243 SNPs with a minor allele count (MAC) of at least 5 used for subsequent analyses. The IMPUTE Info Score takes into account genotype uncertainty.[239]

3.2.1.4 mRNA Expression Data

The Oxford Genomics Centre carried out mRNA expression analysis on the Affymetrix human GeneChip HGU133 Plus 2.0 array (Affymetrix, USA), covering the transcription level of 17,209 Ensembl-annotated genes in 31 abdominal SAT samples. Probes were grouped probes into probe sets corresponding to Ensembl annotated genes by Dr George Nicholson using a custom chip definition file as described by Dai et al.[240] This yielded mRNA expression values as the robust multi-array average (RMA) without background correction, which is a summary measure of quantile-normalised, log2-transformed probe intensities.[241]

3.2.2 The MuTHER Replication Study

The primary meQTL results were followed up by a set of 181 subcutaneous adipose samples from independent female individuals.

3.2.2.1 Samples and Phenotypes

All samples were analysed on the Illumina Infinium HumanMethylation27 BeadChip (27k array) array from 181 female individuals from the MuTHER study.[242, 243] The samples were collected from the TwinsUK resource[244], and previously shown to be comparable to population singletons in terms of disease-related and lifestyle characteristics (**Table 2.2**).[245]

Table 3.2: Participant characteristics for the MuTHER study

Participant characteristics	All Samples
Sample size (N)	181
Age (years)	61.1 (7.6)
BMI	26.4 (4.5)
Weight (kg)	68.9 (12.6)
HDL (mmol/l)	1.9 (0.5)
TG (mmol/l)	1.2 (0.6)
Glucose (mmol/l)	5.1 (0.7)

Values indicate mean (SD) for each trait

3.2.2.2 CpG Methylation Data

The methylation analysis was carried out by the Wellcome Trust Sanger Institute, UK. Abdominal SAT samples were randomised and genomic DNA was isolated using the NORGEN purification kit (Norgen Biotek Corporation, Canada) and quantified using picogreen. Methylation analysis was carried out after bisulphite conversion according to the EZ-96 DNA Methylation Kit (Zymo Research) protocol and performed using the 27k array, which covers 27,578 CpG sites. The methylation analysis was carried out using the manufacturer's protocol (Illumina, USA), using an IlluminaHiScan SQ scanner and BeadStudio 3.2 software for intensity value extraction. This yielded a beta value (methylation score) for each probe corresponding to the ratio of the intensity of the methylated bead type to the combined locus intensity and range from 0 to 1 (unmethylated to fully methylated).

The initial quality control was performed by our collaborator Dr Eshwar Meduri and resulted in exclusion of seven samples including one sample with low bisulphite conversion efficiency (i.e. bisulphite control intensity values <4,000) and six based on CpG coverage (requiring at least 95% coverage per sample), using the BeadStudio detection p-value of signal above background. Probes that failed (N = 1459) and were not reported by the BeadStudio software in at least one individual were discarded. A singular value decomposition (SVD) was performed to determine the nature of the largest components of variation in beta values as previously described.[246] By correlating the top ten principal components to known experimental factors it was shown that BeadChip, BS conversion efficiency (as assessed using the built-in BS conversion efficiency controls) and DNA input contributed significantly to the variation in beta levels and were included together with age included as covariates into subsequent analyses outlined in **Figure 2.1.**

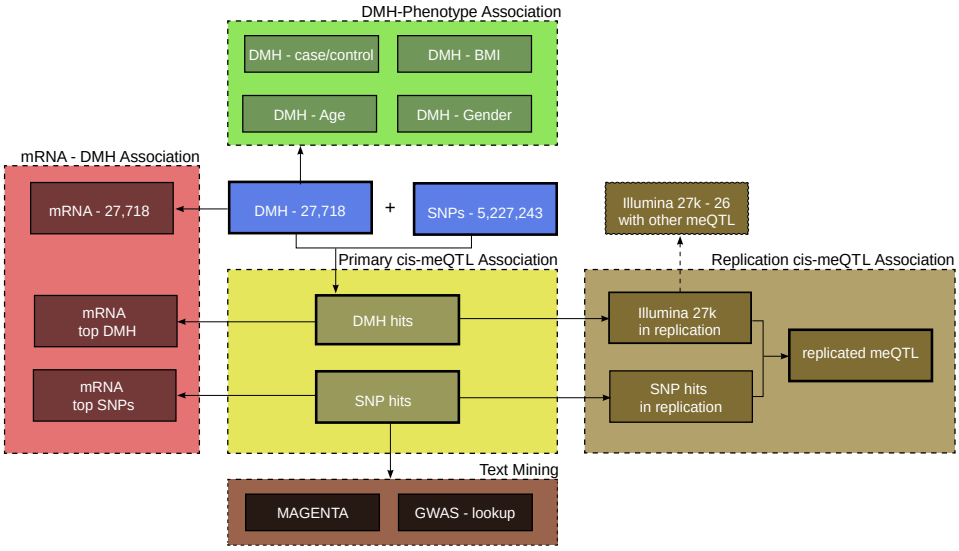


Figure 3.1: Analysis Outline for the MolOBB study. The 27,718 probe sets from the DMH experiment and 5,227,243 SNPs (blue) were used as an input. All analyses were carried out using linear models. The DMH regions were first tested for association with the phenotypes age, gender, BMI and MetSyn (green). Next, SNPs ±500 kb to the DMH probe sets were tested for association (yellow) and carried forward if they reached the <5% FDR threshold. The associated probe sets were then matched to probes ± 2kb in the Illumina 27k platform used for replication and considered significant if the association p<0.05 and were directionally consistent with the discovery. Association for DMH probe sets with mRNA were tested both globally and for the primary hits. This included analysis of top SNPs vs. methylation (red). The hits were also tested for overlap with GWAS hits using MAGENTA and permutation testing (dark brown).

3.2.2.3 Genotype Data

The genotyping of the 181 replication samples was carried out at the Wellcome Trust Sanger Institute, UK as subset of the larger the TwinsUK dataset (N ~ 6000). Quality control was performed by Dr So-Youn Shin and included four Illumina arrays (HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M), for which Intensity values were pooled separately, except for 1M-Duo and 1.2MDuo 1M which were pooled together. The Illumina calling algorithm assigned genotypes if the maximum posterior probability was above 0.95.

Similar exclusion criteria were applied to all datasets. Samples were excluded if (i) sample call rate < 98%, (ii) heterozygosity across all SNPs ≥2 SD from the sample

mean; (iii) evidence of non-European ancestry as assessed by PCA comparison with HapMap3 populations; (iv) observed pairwise IBD probabilities suggestive of sample identity errors. SNPs were excluded if (i) Hardy-Weinberg p-value < 10^{-6}, assessed in a set of unrelated samples; (ii) MAF < 1%, assessed in a set of unrelated samples; (iii) SNP call rate < 97% (SNPs with MAF ≥ 5%) or <99% (for 1% ≤ MAF < 5%). The datasets were then merged after performing pairwise comparison between datasets. Further samples and SNPs were excluded if (i) concordance at duplicate samples <1%; (ii) concordance at duplicate SNPs <1%; (iii) visual inspection of QQ plots for logistic regression applied to all pairwise dataset comparisons; (iv) Hardy-Weinberg p-value < 10^{-6}, assessed in a set of unrelated samples; (v) observed pairwise identity by percent probabilities suggestive of sample identity errors.

Imputation was then performed by Dr So-Youn Shin using the IMPUTE software package (v2.1.1.10) using two reference panels, P0 (1000 Genomes haplotypes, 2011 Phase I (interim) release, EUR) and P1 (the combined HumanHap610k and 1M reduced to 610k SNP content).[238] The SNP positions were mapped from b36 to b37 according the UCSC Feb. 2009 (GRCh37/hg19) assembly and filtered at a MAF > 5% and IMPUTE info value of >0.8 after imputation.

3.2.2.4 Phenotype Association

I applied linear models with the limma R package[214] using the Methylation score[237] of the 27,718 DMH probe sets that passed filtering as the response and case/control status (adjusting for age and gender) as the predictor variable. Similarly, to test for association of either BMI or age with DMH methylation score, models were fitted with methylation as a response adjusting BMI for age and gender, and age for BMI and gender. Lastly, an association of gender with methylation score was tested for the autosomes only, adjusting for BMI and age. A 95% confidence interval was calculated by performing 1000 permutations of the methylation scores for visualisation in quantile-quantile plots.

3.2.3 Cis-meQTL Association – Primary Study

To test for association of methylation with each SNP in ±500 kb region around each of the 29,441 filtered DMH regions, linear additive model was fitted using PLINK. The allelic dosage for each individual i at each SNP marker g was calculated as

$$x_{g,i} = p_{1,i} + 2\,p_{2,i} \quad (1)$$

where $p_{1,i}$ and $p_{1,i}$ are the posterior probabilities for the imputed genotypes AB (heterozygote) and BB (homozygote for the alternative allele), respectively. The genotype AA is omitted from the equation as it represents an allelic dosage of zero. Methylation status at each probe set y was then regressed with the allelic dosage using the following linear model:

$$y = \beta_o + \beta_g x_g + \beta_{age} x_{age} + \beta_{gender} x_{gender} + \beta_{MetSyn} x_{MetSyn} + \varepsilon \quad (2)$$

where the error term is $\varepsilon \sim N(0, \sigma^2)$ for each marker independently. A p-value for $H_0: \beta_g = 0$ was then calculated by using the Wald statistic $\frac{\beta_g^2}{se(\beta_g)^2}$, where $se(\beta_g)^2$ is the standard error in the estimate of β_g. This was then compared to a chi-squared distribution (Wald test). Subsequently, multiple testing was corrected for by applying a false discovery rate (FDR) threshold of 5% (using the *qvalue* package in R).[247]

3.2.4 Cis-meQTL Association – Replication Study

To replicate any *cis*-meQTL findings from the DMH primary cohort, it was first assessed how many probes on the Illumina 27k array were located within a 1 kb distance of DMH probe sets with significant meQTLs in the primary DMH cohort, as there is high correlation of CpG methylation within 1 kb.[163, 179] Of the 149 DMH probe sets with meQTL, 27 had Illumina 27k within 1 kb. This distance was chosen as high correlation of methylation has been reported for this interval.[163, 179] For 19 of these probes there were also SNP genotype data for the lead SNP or a SNP in high linkage disequilibrium (LD, r2 > 0.8) with the lead SNP. To test for replication, a one-

tailed t-test in the direction of the original association, using a Gaussian null distribution as an approximation to the true null distribution (a t-distribution with 176 degrees of freedom).

Briefly, the allelic dosage for each SNP marker in the replication was calculated and the following linear mixed-effect model allowing for twin relatedness was fitted:

$$y = \beta_o + \beta_g x_g + \beta_{age} x_{age} + \beta_{gender} x_{gender} + \beta_{conc} x_{conc} + \beta_{BSC} x_{BSC} + K + \varepsilon \quad (3)$$

where x_{conc} is the concentration after bisulphite conversion, x_{BSC} the bisulphite conversion efficiency measured by the array-internal control and K the kinship matrix (random effect). The above model was fitted by Dr Eshwar Meduri for CpG sites requested, and β_g and $se(\beta_g)$ were provided. A t-statistic calculated as $\frac{\beta_g}{se(\beta_g)}$ was then used in a one-tailed t-test in the direction of the original association, using a Gaussian null distribution as an approximation to the true null distribution (a t-distribution with 176 degrees of freedom)

3.2.5 mRNA Expression Association

To assess the effect of CpG methylation on mRNA expression, each DMH probe set that was significantly associated with a meQTL SNP at the 5% FDR level was matched with the mRNA transcript of which the transcription start site (TSS) was closest in distance. The rationale for this approach is the array design that centred around CpG methylation at gene promoters.[179] A linear model with expression level as the response variable and DMH probe set methylation score as the predictor was fitted, adjusting for age, gender and MetSyn case/control status. A similar analysis was carried out on all 27,718 DMH probe sets.

The meQTL SNPs were then tested for cis-eQTL association with transcripts previously matched to the DMH regions under genetic control. Analogous to the meQTL analysis a linear additive model with allelic dosage as a predictor was fitted,

adjusting for age, gender and case/control status for MetSyn. An FDR threshold of 5% was applied to all analyses.

3.2.6 Text mining and pathway analysis

Firstly, the DAPPLE[248] software was used to investigate protein-protein interactions between the gene products of genes with meQTLs. As an input, the 149 meQTL SNPs and the default settings of 1,000 permutations and a common interactor binding degree cut-off of 2 were used. By default, the gene regulatory region is defined as 50 kb up- and downstream of each transcript.[248]

Secondly, for the gene pathway analysis, each meQTL SNP was assigned to an Ensembl-annotated gene with closest TSS as described in (Chapter 3.2.5p. 53).

Thirdly, to detect whether the meQTL hits in SAT are enriched for associations previously found by GWAS, a permutation test (10,000 permutations) was carried out using a rank-sum statistic, where 149 random SNPs matched for arbitrarily chosen thresholds of MAF ±5%, distance to genes ±500 kb and local recombination rate ±1 cM/Mb in each permutation.

Fourthly, it was tested whether those genes that overlapped cis-meQTL probe sets were statistically over-represented at each term in the GO database[249], relative to the number expected to occur by chance (a hypergeometric null distribution). The resulting p-values were adjusted at the 5% FDR level.

3.3 Results

3.3.1 Patterns of abdominal SAT methylation and validity of DMH Method

Methylation in the MolOBB samples was measured by DMH. This method, based on methylation-specific restriction enzyme digest[235], has been superseded by arrays detecting single CpGs. However, global methylation patterns measured by DMH showed bimodality as expected, with a “hypomethylated mode” at lower and a

"hypermethylated mode" at higher methylation score (**Figure 3.1A**). I found a higher occurrence of hypermethylation in this study compared to what has been previously reported for the Illumina 27k array[179]. This is due to the focus on mostly hypomethylated promoter regions in the Illumina 27k array. In total, 9,253 of the 12,500 genes targeted by the DMH array also are probed by the Illumina 27k array (74.0%)[178, 179], but only 3,629 probed sites directly overlap (13.1%). CpG sites on the X chromosome are intermediately methylated (methylation score 0.3 – 0.7)[179] in females (**Figure 3.1B**), and CpG sites close to the TSS of imprinted genes[250], were approximately intermediately methylated (**Figure 3.1C**). Consistent with earlier studies, I observed lower methylation around transcriptional start sites (**Figure 3.1D**).[163, 179] The consistency with the previously reported patterns of methylation[178, 179] patterns provided validation for the DMH technology.

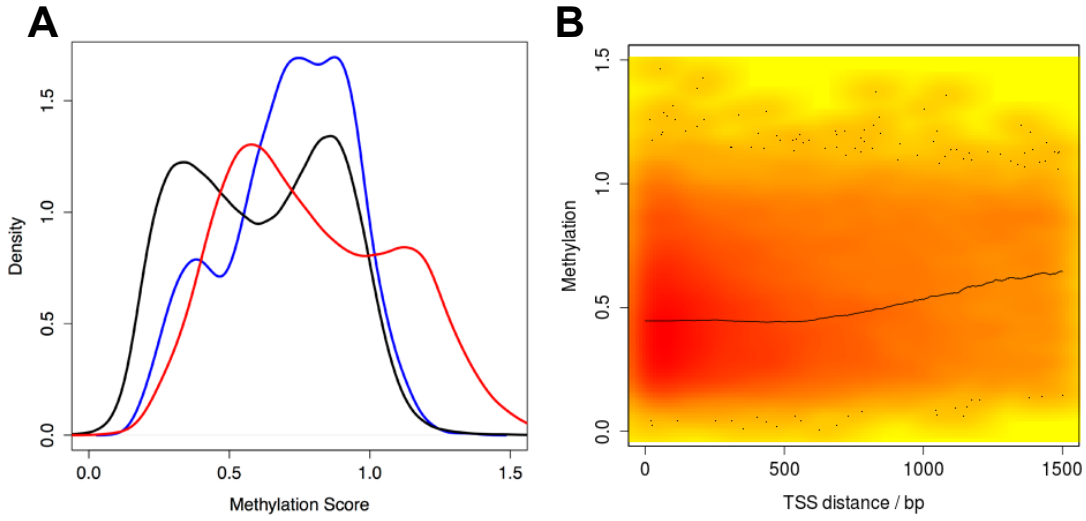


Figure 3.2: Density plots showing methylation patterns in DMH data. A: Across all 27,718 filtered probes, DMH data shows a bimodal pattern consistent with hypo- and hypermethylation (black line). CpG sites on the X chromosome are intermediately methylated in females (red line). CpG sites close to TSS's of Imprinted Genes are intermediately methylated (blue line) B: Lower methylation is observed around TSS's (black line shows the 300bp sliding window median).

3.3.2 Phenotype Association

I then tested whether the DMH methylation score at each probe set associated with MetSyn case/control status or BMI. At the 5% FDR level, no DMH probe sets were significantly associated with MetSyn (**Figure 3.2, black red**) or BMI (**Figure 3.2, blue dots)**. The resulting p-value distribution showed a statistical deflation, but this still remained within a 95% confidence interval generated by 1000 permutations of the methylation scores.

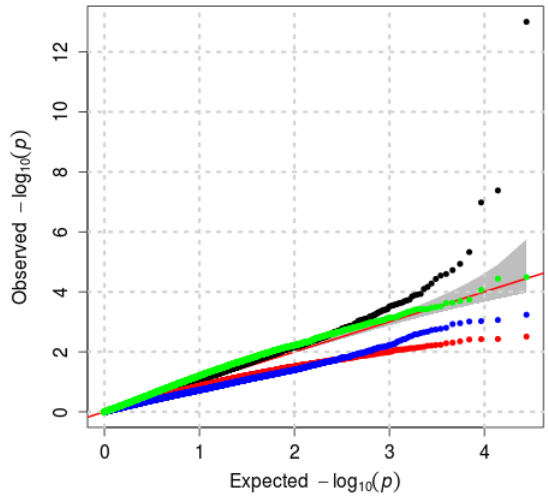


Figure 3.3: Association of 27,718 DMH probe sets with phenotypes. QQ plots show association p-values for MetSyn case/control status (logistic regression, red dots), BMI (linear regression, blue dots), gender (linear regression, chromosomes 1-22 only, black dots) and age (linear regression, green dots). The theoretical 95% confidence interval (CI) is shown in grey.

Repeating the analysis with gender, excluding sex chromosomes, I observed an association signal of gender with methylation score (**Figure 3.2, black dots**). Four autosomal DMH probe sets significantly associated with gender at the 5% FDR level. As with BMI and MetSyn, I detected no association of methylation score with age (**Figure 3.2, green dots**), potentially due to low power (see Chapter 3.3.8 p. 68)

3.3.3 Genome-wide meQTL analysis

I carried out tests for association of the methylation scores with genetic variants in *cis* to each DMH probe set. The results show enrichment of low nominal p-values for SAT meQTLs (**Figure 3.3A**) and 149 meQTL sites were significant at the 5% FDR level (**Appendix Table 1**). The 149 meQTLs show a median distance of 80.7 kb between the DMH probe set and the lead SNP (**Figure 3.3B**, distance range 1 bp – 499 kb). The range highlights that many associations occur at short distance, which is similar to the SNP- CpG distance of 81 kb reported by Gibbs et al. in brain tissue.[178]

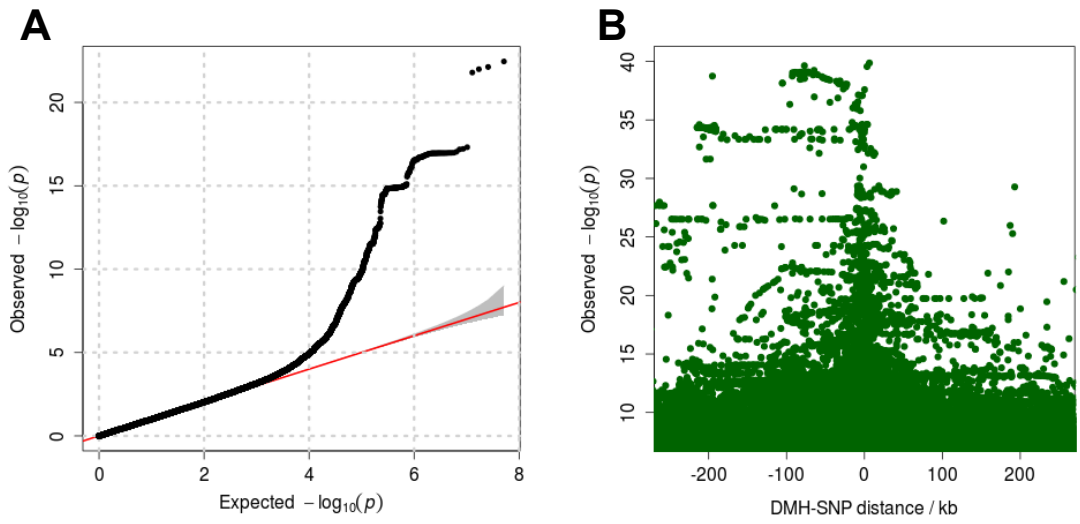


Figure 3.4: Association of 27,718 DMH probe sets with phenotypes. QQ plots show association p-values for MetSyn case/control status (logistic regression, red dots), BMI (linear regression, blue dots), gender (linear regression, chromosomes 1-22 only, black dots) and age (linear regression, green dots). The theoretical 95% confidence interval (CI) is shown in grey.

Of the 149 significant DMH probe sets with meQTL SNPs, 19 have a corresponding Illumina 27k CpG probe (located within 1 kb) and the same SNP, or a proxy SNP ($r2 > 0.8$), available. Five out of 19 tested meQTL SNPs associated (p-value<0.05 in the direction of the association in the primary study) with methylation in the replication (**Table 3.1)**. I found that the replication rate was significantly higher than expected by chance (binomial p-value=2.01×10^{-3}). Regional plots of the SNP

architecture at the five replicated loci are shown in **Figure 3.4** and boxplots showing the correlation between genotypes and methylation are shown in **Figure 3.5**.

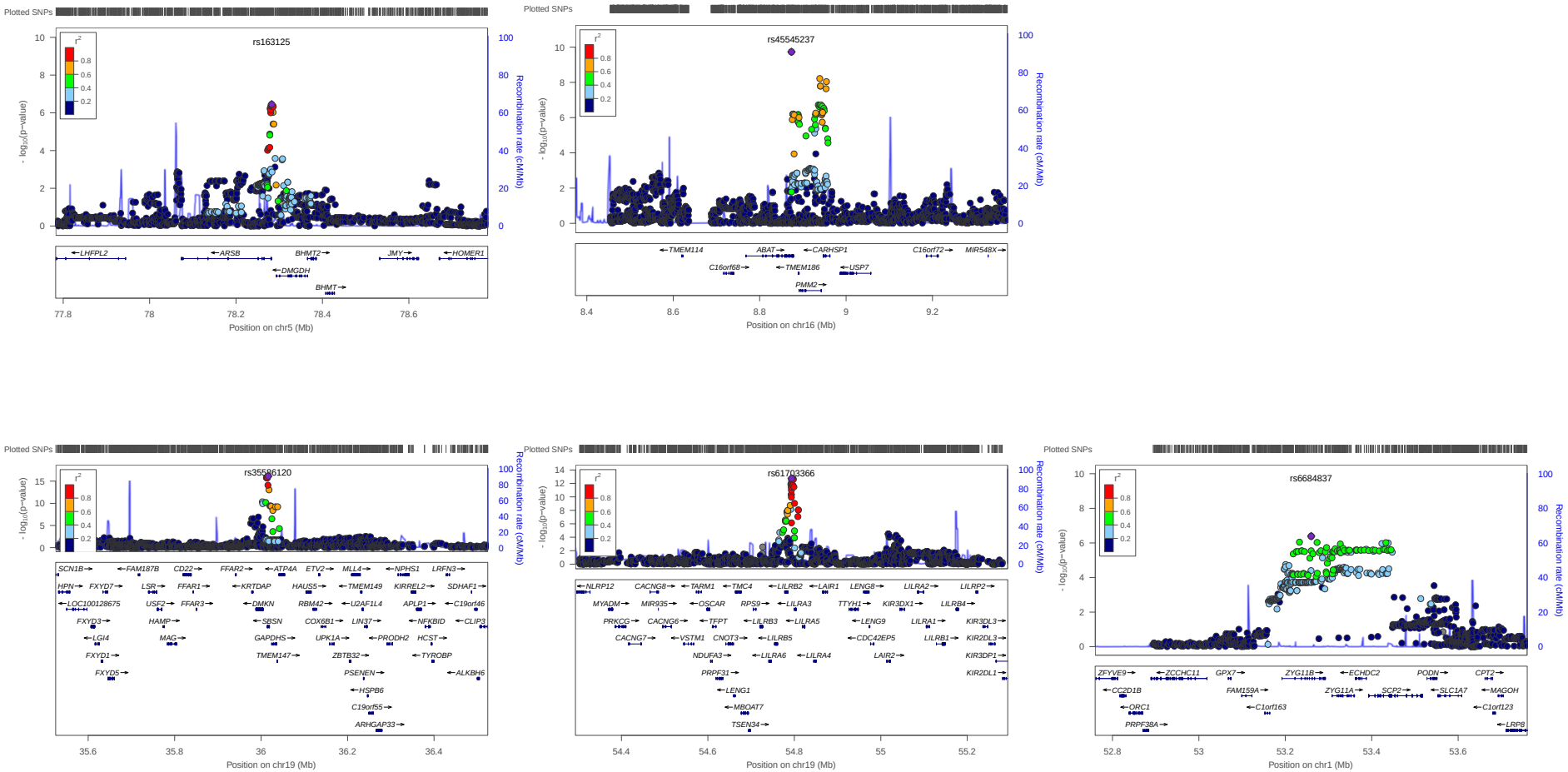


Figure 3.5: Regional plots for 5 replicated meQTL loci. The $-\log_{10}$ *P* values of the 5 replicated (p<0.05 and directionally consistent in the replication study) meQTL loci are plotted against genomic position. The top associated(lowest p-value) SNP in each region is indexed and indicated by a diamond. Estimated recombination rates are shown in blue, and the SNP R^2 is given by colour as shown in the legends (LD data from 1000 Genomes Nov 2010 genotypes). Plots were generated using locuszoom (csg.sph.umich.edu/locuszoom/).

Table 3.3: List of 19 meQTL associated at 5% FDR in the primary study.

Primary Study							Replication Study			
Chr	DMH probe start (b37)	SNP position (b37)	SNP rs ID	Effect	SE	p-value	CpG position (b37)	Effect	SE	One-sided p-value
1	53392620	53259897	rs6684837	-0.0713	0.0113	4.05E-07	53393297	0.0203	0.0031	2.39E-11
1	153631080	153212859	rs1329101	0.0773	0.0133	1.62E-06	153630883	-0.0001	0.0008	4.24E-01
1	153939705	153773093	rs74545118	0.0226	0.0290	4.41E-01	153940420	-0.0001	0.0011	4.68E-01
3	179040338	179247185	rs10513760	-0.3766	0.0634	1.16E-06	179041354	-0.0008	0.0015	7.12E-01
5	78281698	78282670	rs163125	0.1148	0.0181	3.52E-07	78282099	-0.0332	0.0060	1.52E-08
6	166582675	166553633	rs6928596	-0.0627	0.0107	1.47E-06	166582197	-0.0002	0.0009	5.89E-01
7	133812100	133588739	rs6954700	0.0377	0.0058	1.92E-07	133811911	0.0000	0.0005	4.85E-01
10	88282242	88323401	rs2588282	-0.1640	0.0285	1.95E-06	88281865	0.0004	0.0009	3.43E-01
10	30721783	30778184	rs103441	0.1489	0.0253	1.32E-06	30722678	0.0006	0.0011	7.16E-01
14	55907406	55906187	rs12888361	0.2443	0.0269	1.74E-10	55907007	0.0087	0.0032	9.96E-01
15	41625250	41632049	rs7183337	-0.2580	0.0368	5.19E-08	41624793	-0.0012	0.0023	7.04E-01
16	8954289	8873576	rs45545237	-0.7017	0.0775	1.85E-10	8954017	0.0352	0.0070	2.17E-07
17	80708321	80859844	rs7208393	0.1371	0.0170	2.53E-09	80709357	-0.0018	0.0008	9.59E-03
17	55038512	55130175	rs28872233	-0.2041	0.0365	3.28E-06	55038508	0.0002	0.0031	4.74E-01
19	54802668	54793721	rs61703366	0.2151	0.0181	1.78E-13	54803540	-0.0107	0.0058	3.13E-02
19	39390520	38912597	rs4310986	0.0587	0.0100	1.48E-06	39389915	0.0007	0.0010	7.65E-01
19	36024159	36024296	rs74409639	-0.2649	0.0166	4.72E-17	36024439	0.0484	0.0054	2.68E-19
19	44489036	44215450	rs76089163*	-0.2109	0.0430	2.41E-05	44488327	0.0038	0.0029	9.14E-02
22	31364890	31355992	rs142563262	-0.2587	0.0307	1.01E-09	31364948	0.0020	0.0018	1.31E-01

Chr: Chromosome, Effect: beta coefficient for the SNP effect, SE: SE for the SNP effect, p-value: p-value for the SNP effect, One-sided p-value: p-value for the SNP effect in the direction of the original association

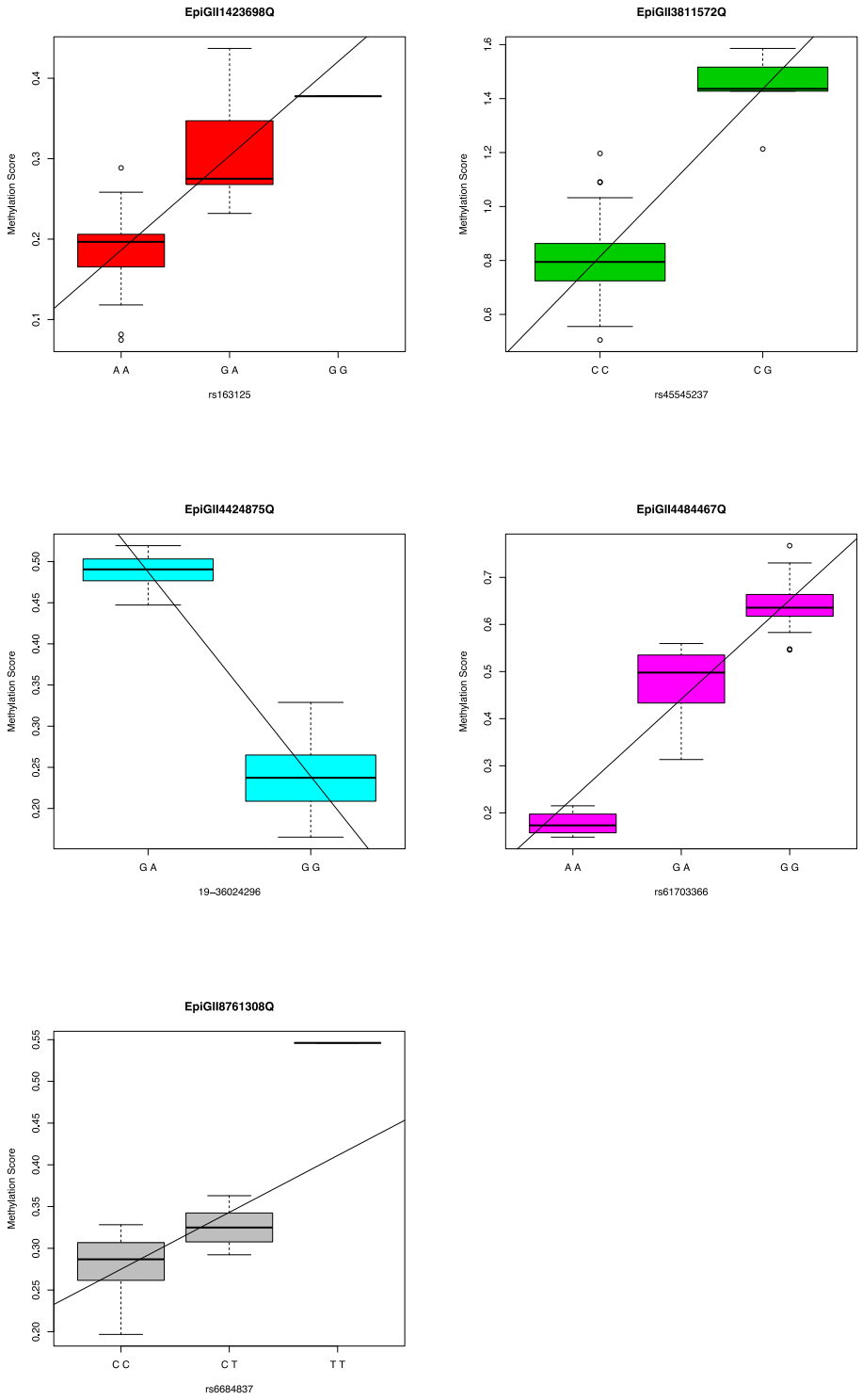


Figure 3.6: Methylation level plotted against genotype per SNP allele at the 5 replicated meQTLs. Boxplots show methylation level plotted against genotype for the 5 replicated ($p<0.05$ and directionally consistent in the replication study) meQTL SNP alleles. The black line indicates the regression line fitted by a linear model excluding covariates at each locus.

3.3.4 Secondary Signals and Conditional Analysis

The 19 Illumina 27k probes were then tested again by Dr Eshwar Meduri for associations with all SNPs within a ±500 kb *cis*-region and β_g and $se(\beta_g)$ of the top associating SNP were forwarded for analysis. Of four probes for which the primary SNP replicated, there were other nearby SNPs in the replication data associated at the 5% FDR level (**Table 3.2**). In one case, the lead meQTL SNP was the same in the primary and replication study. Using conditional analysis, the weaker association signal of the primary and the replication study disappeared when adjusting for the SNP more strongly associated in either study (**Table 3.3**). The interpretation of this is that in some cases there exists an independent stronger *cis*-meQTL in the replication than the one detected in the primary study, and the difference observed is most likely due to heterogeneity between the two sample populations or stochastic effects. For all cases, the top SNP from the replication study also significantly associated with methylation in the primary data (**Table 3.3**), giving further support to the validity of the method.

Table 3.4: Secondary signals in the replication around the 19 meQTLs

CpG ID	Primary top meQTL (SNP1)	P-value (SNP1)	Effect (SE) [SNP1]	Replication top meQTL (SNP2)	P-value (SNP2)	Effect (SE) [SNP2]	Adjusted p-value (SNP2)
cg10272885	1-153212859	8.48E-01	-0.00015 (0.0008)	1-153509890	2.23E-03	0.006(0.0008)	1.00E+00
cg16029636	1-153773093	9.36E-01	-0.00009 (0.0011)	1-153727562	1.38E-02	-0.003(0.0011)	1.00E+00
cg27535305	1-53259897	1.65E-09	0.02026 (0.0031)	1-53392908	4.66E-24	-0.029(0.0031)	5.94E-20
cg18576883	3-179247185	5.77E-01	-0.00082 (0.0015)	3-178687926	1.05E-02	-0.002(0.0015)	1.00E+00
cg24893837	5-78282670	1.41E-07	-0.03316 (0.0060)	5-78279013	3.07E-08	0.036(0.0060)	4.06E-04
cg17188046	6-166553633	8.23E-01	-0.00019 (0.0009)	6-166630197	1.34E-03	-0.004(0.0009)	1.00E+00
cg14706178	7-133588739	9.70E-01	-0.00002 (0.0005)	7-134056838	6.26E-05	-0.004(0.0005)	8.46E-01
cg03807314	10-30778184	5.70E-01	0.00061 (0.0011)	10-30978164	2.09E-03	0.003(0.0011)	1.00E+00
cg19596653	10-88323401	6.87E-01	0.00038 (0.0009)	10-88730762	7.09E-03	-0.006(0.0009)	1.00E+00
cg16249711	14-55906187	8.08E-03	0.00871 (0.0032)	14-55921995	5.32E-04	-0.013(0.0032)	1.00E+00
cg11324740	15-41632049	5.93E-01	-0.00125 (0.0023)	15-41155019	3.36E-04	-0.004(0.0023)	1.00E+00
cg04007936	16-8873576	2.11E-06	0.03519 (0.0070)	16-8951323	1.42E-08	-0.033(0.0070)	2.76E-04
cg07844021	17-55130175	9.48E-01	0.00020 (0.0031)	17-55038490	7.04E-03	0.006(0.0031)	1.00E+00
cg27574244	17-80859844	2.01E-02	-0.00182 (0.0008)	17-80707827	2.05E-03	-0.002(0.0008)	1.00E+00
cg25681177	19-36024296	2.43E-14	0.04838 (0.0054)	19-36024296	2.43E-14	0.048(0.0054)	3.13E-10
cg22814247	19-38912597	4.70E-01	0.00072 (0.0010)	19-38922476	1.48E-02	-0.006(0.0010)	1.00E+00
cg27309253	19-44215450	1.84E-01	0.00383 (0.0029)	19-44488352	1.36E-03	0.005(0.0029)	1.00E+00
cg13733733	19-54793721	7.27E-02	-0.01074 (0.0058)	19-54710673	1.96E-04	-0.020(0.0058)	1.00E+00
cg07894162	22-31355992	2.66E-01	0.00204 (0.0018	22-31265735	8.97E-03	-0.003(0.0018)	1.00E+00

P-value (SNP1): p-value for the effect of SNP1 in the replication samples, Effect(SE): Effect(SE) [SNP1] for the effect of SNP1 in the replication samples, P-value (SNP2): p-value for the effect of SNP1 in the replication samples, Effect(SE) [SNP2]: Effect(SE) for the effect of SNP2 in the replication samples. Adjusted p-value (SNP2): p-value for the effect of SNP1 in the replication samples Bonferroni-corrected for 19 tests

Table 3.5: Conditional analysis around the 19 meQTLs.

DMH ID	Primary top meQTL (SNP1)	Replication top meQTL (SNP2)	$d_{SNP-SNP}$	r^2	D'	d_{DMH-CG}
EpiGII8761308Q	1-53259897	1-53392908	133011	0.37	0.84	0
EpiGII1423698Q	5-78282670	5-78279013	3657	0.89	1.00	206
EpiGII3811572Q	16-8873576	16-8951323	77747	0.31	0.94	272
EpiGII4030067Q	17-80859844	17-80707827	152017	0.23	0.60	-806
EpiGII4424875Q	19-36024296	19-36024296	---	---	---	-7

$d_{SNP-SNP}$: distance between SNP1 and SNP2, r^2: LD correlation coefficient between SNP1 and SNP2, D': LD between SNP1 and SNP2, $d_{SNP-SNP}$: distance between the DMH probe set and CG corresponding probe

Primary Data

DMH ID	SNP1	SNP1 cond SNP2	SNP2	SNP2 cond SNP1
EpiGII8761308Q	4.05E-07	8.07E-04	2.76E-06	5.69E-03
EpiGII1423698Q	3.52E-07	1.00E+00	6.11E-07	1.00E+00
EpiGII3811572Q	1.85E-10	2.19E-08	2.98E-03	4.74E-01
EpiGII4030067Q	2.53E-09	4.75E-06	1.53E-06	2.82E-03
EpiGII4424875Q	4.72E-17	same SNP	same SNP	same SNP

Values represent p-values for the SNP effect in the primary cohort

Replication data

DMH ID	SNP1	SNP1 cond SNP2	SNP2	SNP2 cond SNP1
EpiGII8761308Q	1.65E-09	3.96E-01	4.66E-24	3.13E-16
EpiGII1423698Q	1.41E-07	6.74E-01	3.07E-08	7.17E-02
EpiGII3811572Q	2.11E-06	1.36E-02	1.42E-08	7.19E-05
EpiGII4030067Q	2.01E-02	4.72E-01	2.05E-03	3.16E-02
EpiGII4424875Q	2.43E-14	same SNP	same SNP	same SNP

Values represent p-values for the SNP effect in the primary cohort

3.3.5 Comparison between published meQTL studies

Next, I investigated whether the meQTLs found by DMH in this study are also meQTLs reported in other tissues (i.e. LCLs or brain tissue).[178, 179] Due to the limited overlap, only 27 of the 149 probe set - SNP tests could be compared. There was no overlap between the meQTL associations found and the results from the HapMap LCLs[179], but amongst the significant meQTLs reported by Gibbs et al. in brain tissue[178], there are two of 887 probes (cg24893837 [*ARSB*], cg25681177 [*GAPDHS*]) that associate with the same SNP in the DMH data (FDR <5%). An additional probe (cg04007936 [*CARHSP1*]) showed association with a SNP that was independent of the lead SNP in this study (R^2>0.1). Association data for the three loci from the Gibbs et al. study are shown in **Table 3.4**.

Table 3.6: Association data in brain tissues for the 3 loci overlapping with the Gibbs et al. meQTLs.

Tiss	CpG ID	Ch	CG Pos	Loc	SNP	SNP Pos	p-value
crbl	cg24893837	5	78317805	*ARSB*	rs163125	78318426	6.22E-22
fctx	cg24893837	5	78317805	*ARSB*	rs163125	78318426	7.29E-18
pon	cg24893837	5	78317805	*ARSB*	rs163125	78318426	1.44E-13
tctx	cg24893837	5	78317805	*ARSB*	rs163125	78318426	1.06E-15

Tiss: Tissue, Ch: Chromosome, Loc: Locus, Pos: Position, crbl: cerebellum, fctx: frontal cortex: pon: pons, tctx: tectoral cortex

3.3.6 Association of mRNA expression with methylation

Association of DMH methylation scores with *cis*-mRNA expression is a potentially interesting downstream mechanism of the meQTL SNP.[178, 179] When taking all 27,718 DMH probe sets into account, no associations of methylation with expression of the transcript with the closest TSS remained significant at the 5% FDR level (**Figure 3.6A, black dots**). However, when I only considered those 149 probe sets for which meQTLs existed, there is an association signal between methylation score and mRNA expression (**Figure 3.6A, blue dots**). Similar to the SNP-methylation associations, highly correlated associations between DMH probe sets and nearby transcripts tended to occur at short distances (**Figure 3.6A, green dots**). These results showed that two out of 149 mRNA transcripts were significantly associated with DMH methylation score at the 5% FDR level: EpiGII2424610Q (*TNFRSF11B*, Tumour necrosis factor receptor superfamily member 11B), and EpiGII8555714Q (*GOT1*, Aspartate aminotransferase). The distances between the TSS and the correlating DMH probe set were 324 kb and 13 kb, and the correlation between DMH methylation and mRNA expression levels was negative in both cases. To test if the 149 meQTL SNPs were also eQTL SNPs, I tested association of 149 lead meQTL SNP genotypes with mRNA expression (**Figure 3.6B**). No associations were significant at the 5% FDR level. However, due to the low sample size and limited number of tests, association of meQTL SNPs with expression cannot be ruled out.

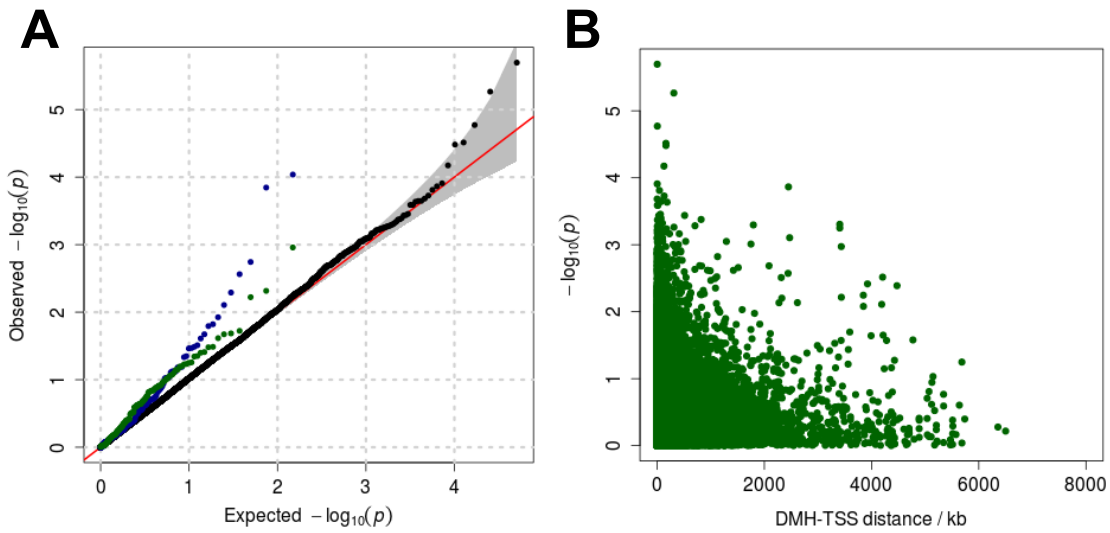


Figure 3.7: Analysis of DMH probes / SNPs for association with expression levels of downstream transcripts. A: QQ plots show p-values for association tests of 27,718 DMH probes (black dots), 149 DMH probes associated with SNPs at the 5% FDR level (blue dots), 149 top meQTL SNPs (green dots). B: Scatterplot of –log10(*p*) values of association with transcripts plotted against the distance between the transcripts and DMH probe sets. The theoretical 95% CI is shown in grey.

3.3.7 Gene Enrichment Analysis

To find potential shared pathways by which the meQTLs act, I carried out an interaction analysis between protein products of genes close to the meQTL SNPs. There was no enrichment of direct (p=0.93) or indirect (p=0.43) interactions. Using a permutation approach with matched SNPs (see Materials and Methods), enrichment for loci previous implied by GWAS was investigated amongst the 149 meQTL SNPs, yielding mostly negative results however: (i) BMI[86] (p = 0.69), (ii) WHR adjusted for BMI[87] (p = 0.71), (iii) TC (p = 0. 87), (iv) TG (p = 0.04), (v) HDL[251] (p = 0.77), (vi) and LDL[251] (p = 0.32)).

A similar analysis was then carried out using MAGENTA for the six phenotypes above, which uses the full set of GWAS results for each phenotype. In the analysis, each DMH probe set is assigned a gene, similar to the mRNA transcript assignment (see above). Of these, HDL associated genes (**Table 3.5**) had significant enrichment in the 149 genes tested ($p = 7.40 \times 10^{-3}$).

Lastly, a GO pathway analysis was performed to identify any biological pathways not captured by the above tests in which the meQTLs could potentially be involved. However, no significantly enriched GO terms were found with the FDR controlled at 5%.

Table 3.7: Enrichment results from the MAGENTA GSEA analysis.

Trait	Expected Genes	Observed Genes	GSEA p-value
BMI	2	1	8.45E-01
HDL	2	6	7.40E-03
LDL	2	2	5.83E-01
TC	2	3	2.68E-01
TG	2	2	5.40E-01
WHR	2	3	2.97E-01

Trait: Input GWAS results used in analysis. Expected Genes: Expected number of genes below the p<0.05 threshold based on 10,000 permutations. Observed genes: number of genes below the p<0.05 threshold based on 149 meQTLs. GSEA p-value: p-value for enrichment at the p<0.05 threshold.

3.3.8 Power Calculations

The effect sizes that were detectable with the sample size in this study, N = 38, and a significance level specified in the multiple-testing setting were calculated. A conservative Bonferroni correction strategy was used, meaning that if 27,578 DMH probe sets are tested the adjusted significance level is $\alpha = 0.05 / 27{,}578$, i.e. $\alpha \sim 1e\text{-}6$ (**Figure 3.7A**). At a fixed α of $1x10^{-6}$ and 80% power, MetSyn case/control associations down to a standardized effect size of 2.16 can be detected. Further, the detectable effect size in units of methylation score was estimated for each probe set, since the standardized effect size is scale-free and the within-group standard deviation (SD) can vary across probe sets. In this data, there is thus 80% to detect effect sizes of 0.05 differential methylation at 14 of 27,718 probe sets and 0.10 methylation score difference at 4123 of 27,718 probe sets, where 0.05 and 0.10 here refer to difference in methylation score (not statistical significance level).

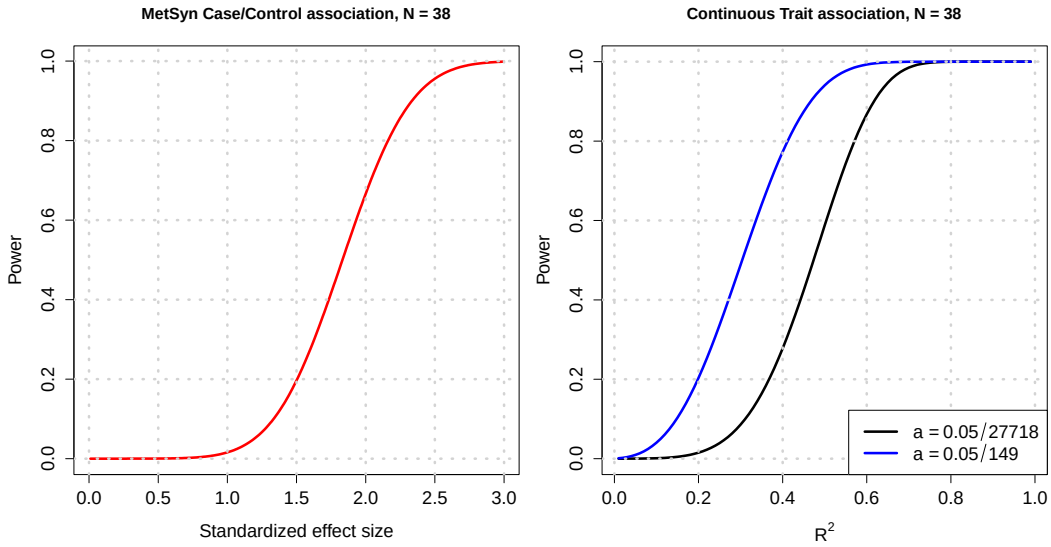


Figure 3.8: Power calculations A: Power curve of power against standardised effect size, at fixed N=38 and $\alpha=10^{-6}$. B: Power calculation curve of power against percent variance explained, at fixed N=38 and two fixed alpha levels ($\alpha=0.05/27,718$ and $\alpha=0.05/149$). The different α levels and curves arise from differences in multiple testing both for all DMH probes (27,718) and significant meQTLs (149).

Using the equivalent values of $\alpha=1x10^{-6}$ and 80% power there is power to detect methylation-BMI associations explaining 57% of variation in BMI (**Figure 3.7B**). Thus, a lack of power and subtle MetSyn or BMI effects in this relatively small study can explain the lack of detection of significant associations.

Using a similar assumption as above, and adjusting for 149 tests, there is 80% power to detect SNP-mRNA associations explaining 41% of variation in expression (**Figure 3.7B**). Again, the detection of associations of this magnitude is limited by the small sample size.

3.4 Discussion

Epigenetic marks, such as DNA methylation, are generally not inherited across generations[171], but DNA sequence variants that associate with methylation (meQTL) have been found throughout the genome for a number of tissues[178, 179, 227, 228]. Not only do genetic effects on methylation provide a mechanism by which methylation patterns can be transmitted across generations, but also provide a biological mechanism for disease-associated SNPs.[179]

In this chapter, the main objective was meQTL analysis. In addition, I also investigated whether DMH methylation associated with MetSyn status, BMI, age or gender. However, no associations of DMH methylation with MetSyn case control status or BMI were found. Changes in DNA methylation associated with BMI have been previously reported[211], but these effects might not have been observed in this study due to lower sample size, and discovery of BMI methylation loci be revisited in Chapter 5. Consistent with some part of the literature[163, 179, 252] an association between methylation and gender was observed, but association between methylation and age was found. This is potentially due to the low power to detect differences and narrow age distribution (SD = 5 years, Table 3.1). As the results indicate, there is more power to detect significant associations due to larger effect sizes of such interactions on the molecular or biochemical level.

The results of the meQTL study indicate that there are a number of *cis*-meQTLs in abdominal SAT, and five out of those DMH probe sets that reached significance were also found to be significantly associated in the same direction in the replication study. Each top *cis*-meQTL SNP, within a ±500 kb region, also is a meQTL in the other data set, which shows consistency of associations of methylation with SNPs in these five regions are consistent throughout the datasets. Thus, the replication study carried out on the Illumina 27k array also provided validation of the DMH method used in the primary study. Moreover, the differences between the two platforms may have affected the replication rate in three ways: (i) by targeting of single vs. multiple CpGs, (ii) by imperfect correlation between the DMH probe sets and Illumina 27k probes (within 1 kb) and (iii) by technical differences between the two assay and laboratory techniques.

Additionally, different genotyping arrays were used between the studies. Stringent quality control measures were used before and after imputation, filtering out low allele frequency and low quality variants. As it was shown that for common variants

(MAF >5%) the imputation accuracy performs similarly well[253], the association results are as robust as they can be to genotyping and imputation errors. DNA methylation associates with genetic variants consistently across the two studies despite differences in assay technologies. This suggests that there is indeed genetic control of methylation levels at specific loci in SAT.

The 149 meQTL SNPs did not associate with expression of *cis*-mRNA transcripts, potentially also due to power issues. Moreover, the association test might also be affected by the fact that methylation and expression analysis was carried out from different aliquots. The association of meQTLs with gene expression has been debated in the past literature. A similar observation that many meQTLs do not influence mRNA expression was made by Gibbs et al.[178], while Bell et al.[179], observe a significant enrichment for eQTLs in the meQTL results. The largest study so far, by Grundberg et al., found that only 6% of meQTL are also eQTL.[232] This controversy highlights that fact that there is still need for larger future studies to investigate the effects of DNA methylation and meQTLs. The meQTL results show a significant enrichment for meQTLs found in brain tissue[178], but not for those in HapMap LCLs.[179] This may be caused by limited overlap between the DMH probe sets and the Illumina 27k microarray, tissue-specific differences, or low power in the initial studies. Future meQTL studies with larger sample size could reveal the true degree of meQTL tissue specificity.

Two *cis*-mRNA's significantly associated with meQTL DMH probe sets. The genes encoding these transcripts were previously implicated in T2D and MetSyn. *TNFRSF11B* (also known as *RANKL*), was previously characterized as an extracellular negative regulator of osteoprotegrin, which acts as a decoy receptor when secreted[254]. A number of studies have found that both osteoprotegrin and this TNF-superfamily protein have elevated serum levels in T2D patients[255, 256], and MetSyn[257]. The protein product is normally secreted by osteoblasts [254]. *GOT1*, also

known as *AST1*, encodes a liver transaminase that plays a role in amino acid metabolism, the urea cycle and the Krebs cycle.[258] The *GOT1* gene promoter has been shown to be regulated by glucocorticoids, cAMP and insulin.[259] A role of these genes in SAT has not been hypothesized or investigated previously.

This study showed, for the first time, that meQTLs are present in adipose tissue. This direct genetic influence on DNA methylation can also indicate an indirect influence on the general molecular phenotype of adipose tissue. Defining the genetic influence on both gene expression and CpG methylation in abdominal adipose tissue is a step forward in understanding molecular pathways associated with obesity and characterising this type of tissue.

Chapter 4 Illumina 450k Method Assessment and Development

4.1 Introduction

Similar to their genetic counterpart, EWAS utilise data from large population or case-control studies to scan for phenotype associations at a large number of epigenetic markers, such as those for DNA methylation.[192] The emerging interest in these studies made the development of high-throughput methods for the measurement of DNA methylation necessary. There are a number of both epigenome-wide and targeted platforms available, which complement each other in throughput, cost, and coverage.[260, 261] However, with each method there are both technical and intrinsic biological issues, which highlight the need for method development in this area. I will describe these challenges in the following paragraphs.

4.1.1 Tissue and Cell Specificity of Methylation

Recent EWAS studies for obesity and T2D, as well other common diseases, have measured methylation in peripheral blood.[213, 262] While there is criticism that this might not represent methylation in tissues where disease pathways are active[263], whole blood can easily be collected and stored for a large number of subjects. Moreover, blood could be a proxy for epigenetic marks in other tissues, as there are reports that methylation at certain sites in blood correlates with those in other tissues.[264, 265] This correlation depends on which part in development methylation has occurred and there is particularly higher agreement between epigenetic marks established in early life.[266, 267] In terms of usefulness for diagnostics and biomarker discovery, a highly accessible tissue will be key.

However, blood is a heterogeneous tissue with various cell types, such as immunologically active white blood cells (WBCs). As these might be present in

different proportions in cases and controls due to inflammation both in T2D and obesity[268, 269], an analysis of differential methylation might be confounded and give rise to spurious results. Since measurements of the individual white blood cell proportions are not always available, Houseman et al. proposed a method for inferring proportions of white blood cell subsets in whole blood samples.[270] The method utilises methylation signatures at informative markers from a reference of purified cells to infer the proportions in whole blood.[270]

4.1.2 Temporal dependence

The status of an epigenetic marker is spatially and temporally dependent, analogous to RNA transcription levels. The time-course over which changes occur has not yet been clearly defined, but studies have shown that changes occur even after a short exercise intervention of six months.[271] This plasticity of methylation means that reverse causation, where the methylation is driven by the phenotype of interest and not causal, will be difficult to distinguish from true causation. Another problem any cross-sectional EWAS would face is determining how stochastic random changes in methylation over time affect single markers used for detection.[211]

4.1.3 Assays for Epigenetics

Over the last 10 years, many different assays for the measurement of DNA methylation have been developed, first starting with those based on immunoprecipitation or enzymatic digestion of CpG containing regions.[260] Whilst these methods are appropriate for measuring the average methylation over genomic regions, single site-resolution methods are based on modifying the extracted genomic DNA by sodium bisulphite conversion. This method converts any unmethylated cytosines via catalytic deamination to uracil, which is in turn amplified as a tyrosine during the polymerase chain reaction (PCR, **Figure 4.1**).[260, 261]

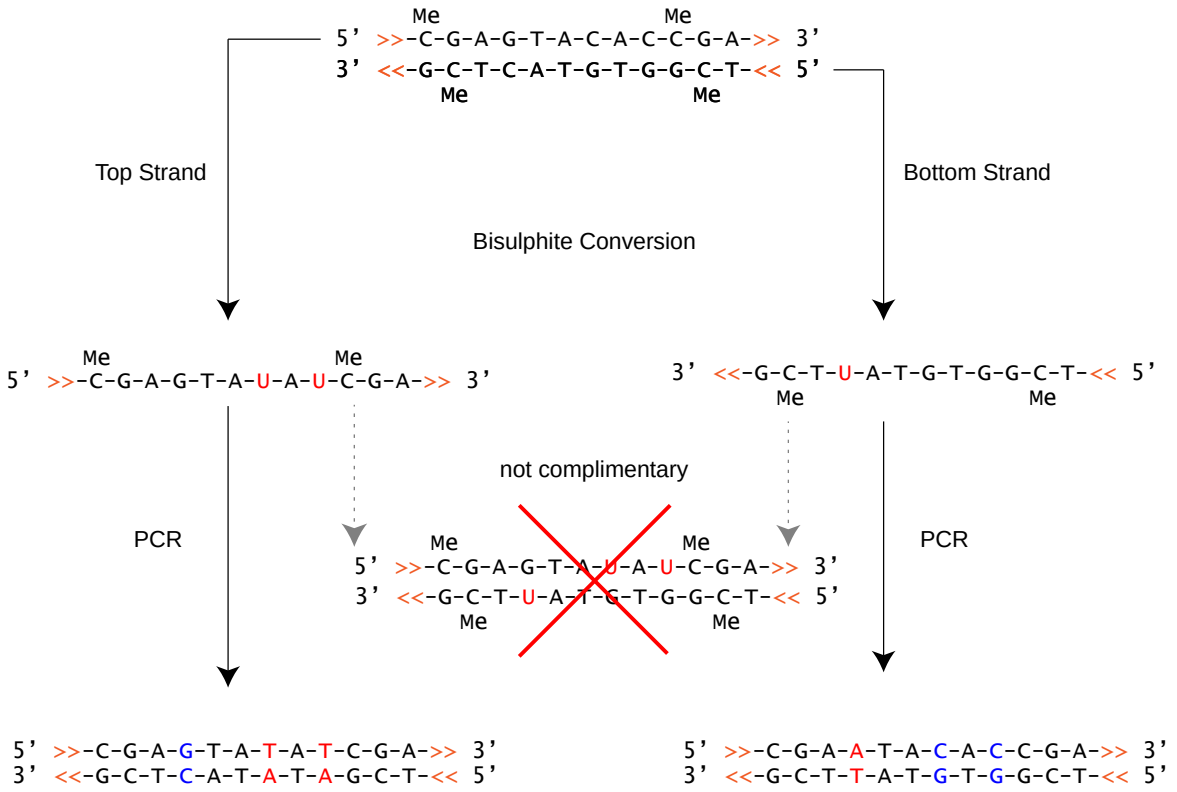


Figure 4.1: Bisulphite conversion of genomic DNA and subsequent PCR amplification. Bisulphite conversion converts unmethylated cytosines (C) to uracil (U), while methylated cytosine residues are resistant to bisulphite conversion. Bisulphite-converted DNA is not complimentary anymore due to U/G mismatches. The resulting PCR gives rise to two PCR products and up to four DNA strands. Me: 5-methylcytos

This essentially means that a cytosine/tyrosine SNP is introduced, which can then be measured by modified conventional genomics methods, such as sequencing (for example pyrosequencing or next-generation sequencing) or microarray hybridisation. The latter has been popularised by the introduction of the Illumina Infinium BeadChip microarray technology, first in form of the 27k[272] and later 450k[191] array, on which this chapter will focus.

The design of the 27k mostly centred on gene promoters and CpG islands[272], which is useful for cancer studies, where epigenetic patterns are dramatically altered.[273] The discovery that there is higher inter-individual variation in other regions such as CpG shores around CpG islands[274], however, led a broader design of the higher-density 450k array, which can give a less biased insight into individual methylomes (**Figure 4.2**).

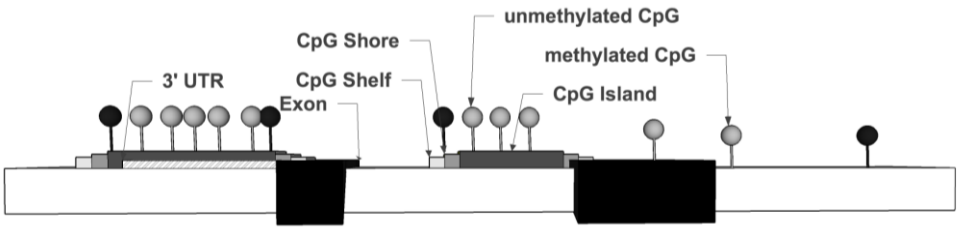


Figure 4.2: Schematic diagram of CpGs relative to the nearest CpG island and gene features. Regions of high CpG density, known as CpG islands (dark grey) are are flanked by CpG shores (medium grey) at ± 2 kb, which in turn are flanked by CpG shelves (light grey) at ±2 kb. CpG islands are commonly found near 3' UTRs (diagonal cross-hatch), but also within genes. Within CpG islands, CpG sites are mostly unmethylated (grey circles), whereas outside CpGs island they are predominanty methylated (black circles).

Both the 27k and the 450k arrays rely on the Illumina Infinium BeadChip technology, which is based on detection of nucleotides by a single-base pair extension after hybridisation of a 50-mer probe sequence to the bisulphite-converted DNA.[272] This is followed by a single base pair extension with fluorescence-labelled nucleotides, which gives rise to signals in two colour channels.[191] The assay chemistry features

two types of probes (**Figure 4.3**): Type I probes, which are the only type used on the 27k array, measure the DNA base next to the CpG site targeted, and thus require two probes per site (one for the methylated and one for the unmethylated CpG site). As the following base is the same for both probes (**Figure 4.3**), the signal is measured in the same channel. The 450k array, on the other hand features both Type I and Type II probes. The Type II probes specifically designed for the 450k array measure both the unmethylated and methylated signal with the same probe, but since the single-base extension is carried out on the G base in the CpG site (**Figure 4.3**), the unmethylated signal always arises from the red channel, while the methylated channel is based on the green channel. For both probe type, the methylation score ("beta value") is calculated as the methylated signal divided by the sum of the methylated and unmethylated signal (see Chapter 2).

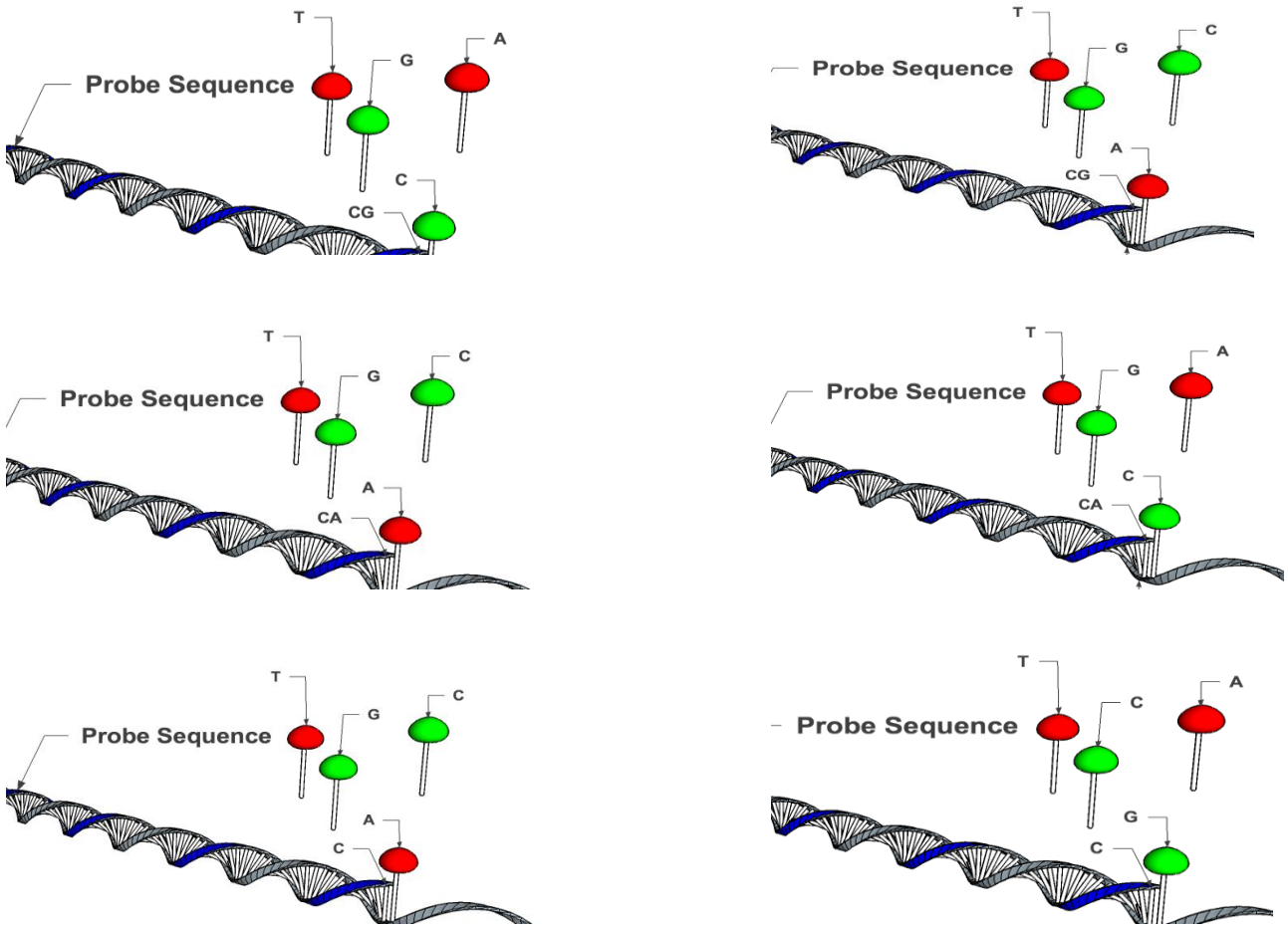


Figure 4.3: The Illumina 450k assay. A: Illumina Infinium Type I probes, B: Illumina Infinium Type II probes. Nucleotides adenosine (A) and tyrosine (T) are detected in the red channel, whereas C and G are detected in the green channel. For Type I probes, methylated and unmethylated signals are measured with two different probe types, as the base following the CpG site is measured. With two colour channels, this gives rise to 4 categories. Type II probes measure both unmethylated and methylated signals using the same probe, giving rise to 2 categories. This leads to 6 possible combinations of probe type and colour channel.

The positioning of the probes on the 450k array has been designed to focus both around genes, as well as CpG Islands, Shores and Shelves (**Table 4.1**). Given that there are over 28 million CpG sites in the genome[275], the ~450,000 probes on the array would thus only represent ~1.5% of the methylation sites. Furthermore, as a panel of experts[191] selected the array content based on the knowledge of methylation variability at the time, array bias can be introduced into downstream analyses.

Table 4.1: Coverage of the 450k array with respect to genes and CpG Islands.

Genic feature type	Genes mapped	Percent genes covered	Average number of loci on array	Transcript type
TS200	14,895	0.79	2.56	coding
TS1500	17,820	0.94	3.41	coding
5'UTR	13,865	0.78	3.34	coding
1stExon	15,127	0.8	1.62	coding
3'UTR	13,042	0.72	1.02	coding
GeneBody	17,071	0.97	8.97	coding
TSS200	1967	0.65	1.84	noncoding
TSS1500	2672	0.88	2.92	noncoding
GeneBody	2345	0.77	5.34	noncoding

Island feature type	Islands mapped	Percent islands covered	Average number of loci on array
Island	26,153	0.94	5.08
N_Shore	25,770	0.93	2.74
S_Shore	25,614	0.92	2.66
N_Shelf	23,896	0.86	1.97
S_Shelf	23,968	0.86	1.94

TS200: within 200 bp of a TSS, TS1500: within 1500 bp of a TSS. N: North Shore/Shelf (upstream of a CpG Island) S: North Shore/Shelf (downstream of a CpG Island). Information based on the 450k annotation provided by Illumina (http://res.illumina.com/documents/products/datasheets/datasheet_humanmethylation450.pdf).[191]

With respect to the detection of methylation differences using the 450k array, methods can be classified into single-marker analyses, which identify differentially methylated positions (DMPs) and regional analyses, which identify differentially methylated (genomic) regions (DMRs). While recently more and more algorithms are being developed to detect DMRs[276], scans for DMPs have the advantage of not being affected by the sparsity and irregularity of the CpG sites probed on the 450k array. For example, direct translation of methods developed for tiling arrays to the 450k array revealed only useful results for 20% of the array content.[277] This has been improved by other approaches that establish methylation patterns over regions by using functional annotations to group probes in the primary analysis.[278] An alternative approach could be to follow up single-marker tests in regions of interest with targeted bisulphite sequencing of multiple nearby CpGs.

The 450k array technology has been validated[279] and used for initial EWAS, including those for traits related to obesity and T2D.[202, 213] However, it was noted after release of the platform that there was need for statistical method development, as there are many technical confounders. These include batch effects and the observed non-uniform behaviour of methylation scores across probe types.[217] This led to the need for unified normalisation and common references for the Infinium methylation arrays.

For all of the above issues, multiple methods have been proposed for the analysis of complex data generated by the 450k array. These include methods for processing raw data[215, 219, 280, 281], normalisation across arrays[215, 216, 219, 278], probe bias correction[217, 218] and batch effect correction.[282, 283] However, a consensus pipeline specifically for the efficient analysis of large scale EWAS data has not been developed and there is need for systematic evaluation of the existing methods.

4.1.4 Questions

I have addressed the methodological issues outlined above, as well as others, during the course of my thesis. Some of these were addressed in a Pilot Study of 90 samples from 45 South Asian men from the LOLIPOP study measured at different time points, designed to ask the following questions:

a. Are there global effects of T2D treatment on methylation?
b. Does methylation change with aging, and does this vary between cases and controls?
c. What are confounding factors (such as white blood cell subsets) that need to be corrected for?

Based on the results from the pilot study, a main study was implemented that measured methylation data of ~2,600 samples from the LOLIPOP study as part of the EpiMigrant study, including 36 samples measured in duplicate. I used the data to evaluate a number of analytical questions:

d. What are appropriate QC parameter thresholds and which metrics can be used to assess this?
e. Which normalisation method performs best?
f. Which probes on the 450k array need to be removed from the analysis due to inherent problems?
g. Do the samples cluster by biological phenotypes and are there outliers?
h. What is the extent of global correlation structure and how can the null-hypothesis can be recovered for association testing?
i. What are the factors driving the correlation structure and batch effects?
j. Can adjusting for these confounders recover the null hypothesis?
k. Does adjustment for confounders influence local correlation patterns?

Lastly, insights from the above analyses were used to assemble a workflow and analytical pipeline and I tested its performance against other published methods through spike-in simulations and computational scale up.

4.2 The Pilot Study

To investigate several effects of study design and experimental factors, I analysed 450k methylation data from a Pilot Study (**Figure 4.4**). These comprised samples from 45 South Asian men matched for age selected at baseline (50.9 ± 9.1 years) from the LOLIPOP study. At baseline, all individuals were normal glucose tolerant (NGT). Follow-up samples were selected such that 7.0 ± 0.9 years later, 15 subjects remained NGT, while 30 developed T2D. Of the 30 people with incident T2D, 15 were previously undiagnosed and untreated for T2D at the time, while 15 had been clinically diagnosed and were on treatment. This design allowed me to answer a number of preliminary questions.

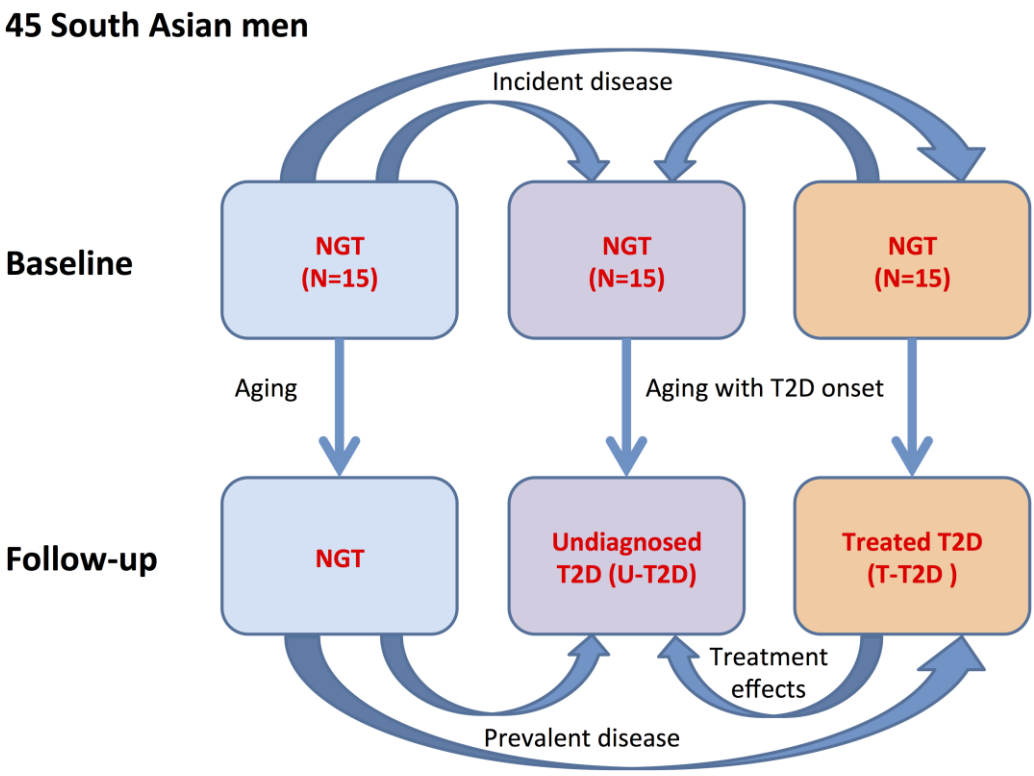


Figure 4.4: Design of the EpiMigrant Pilot Study. Samples were chosen so that at follow-up 15 subject were NGT, 15 were U-T2D and T-T2D, with all subjects classified as NGT at baseline. This study design allows for a number of tests: First, at baseline, association with incident T2D can be tested (top). Similarly, at follow-up, association with prevalent T2D and treatment effects can be tested (bottom). Within and between the groups, the effect of aging can be tested by comparing methylation at follow-up with baseline.

4.2.1 Global effects of T2D or T2D treatment on methylation

If the onset of T2D is associated with changes in methylation, efforts to identify predictors of future T2D may be confounded by use of retrospective cases. This may be further complicated if T2D treatment has an effect on methylation. In this section, I investigated whether there are pronounced effects of treatment within the Pilot Study. As the sample size is low, my analyses focussed on global effects across all markers and enrichment, rather than detecting single markers.

A linear model was fitted for all CpG sites and group comparisons (NGT vs. U-T2D, NGT vs. T-T2D, U-T2D vs. T-T2D). Global methylation differences (across all markers) between those individuals that develop T2D and the controls were

investigated using the test statistic inflation (λ, see 2.5.1 p. 40) in the baseline samples, which may affect single-marker detection.

Global differences in methylation beta values were detected between the incident U-T2D group and NGT controls (**Figure 4.5A**), as well as the incident T-T2D group (**Figure 4.5B**). In contrast betas were similar between U-T2D and T-T2D (λ=0.98, **Figure 4.5C**). The comparison of mean methylation values between NGT and the combined T2D groups (**Figure 4.5D**) shows that most methylation markers either stay in the methylated or unmethylated regions for both cases and controls, with small systematic differences between the groups.

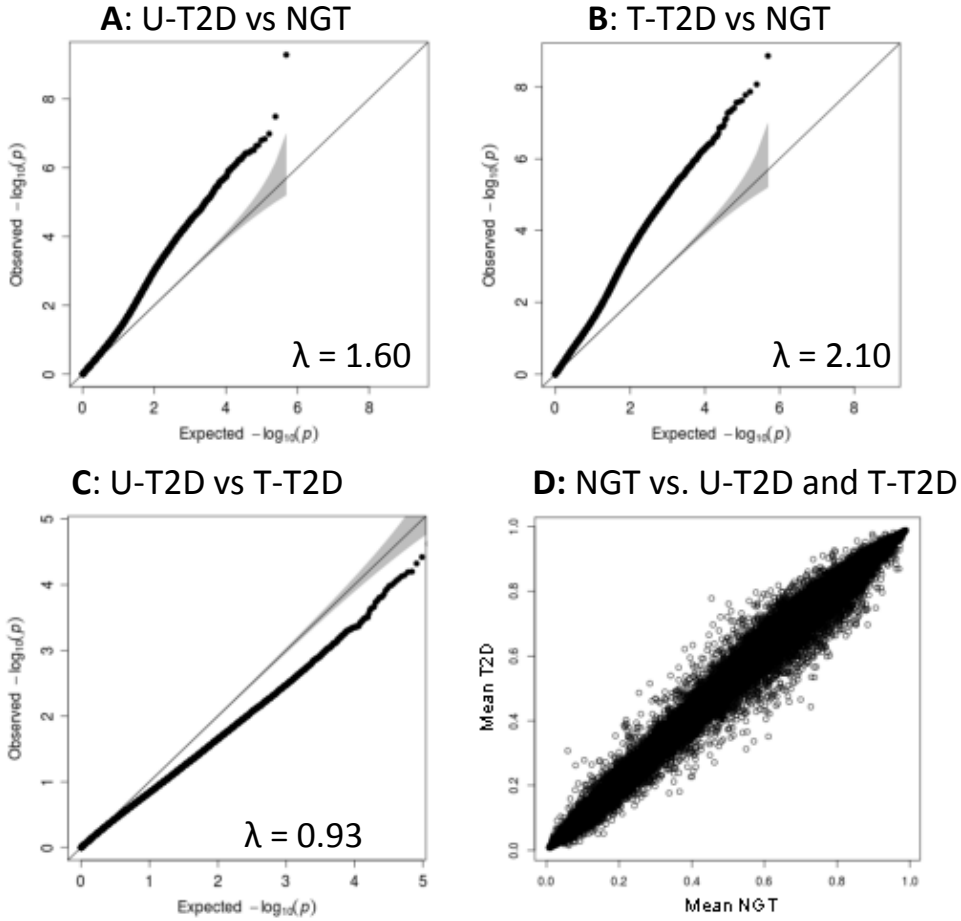


Figure 4.5: Association of epigenetic markers with incident T2D in the EpiMigrant Pilot Study. The QQ plots for p-values from linear models of association between methylation against and status (NGT, future U-T2D and future T-T2D) are shown for all three groups, where N=15 for each group. Genomic inflation factors (λ) for all association tests are given in the right corner. A: NGT vs. U-T2D B: NGT vs. T-2D, and C: U-T2D vs. T-2D. D: Scatterplot showing the correlation between mean T2D case and control beta values at each marker.

The results indicate that there are systematic, epigenome-wide associations of methylation with T2D. These mostly represent small changes in mean methylation. This means that small effect sizes can be expected in EWAS for common diseases such as obesity and T2D.

4.2.2 Methylation change with aging between cases and controls

Multiple studies have suggested that methylation in the same individual can change over time.[211, 267, 284, 285] For the EpiMigrant study, the question arises whether methylation in blood changes between the two time points at which the samples were collected. If so, the predictive value of methylation might be limited. Furthermore, if these changes are different between cases and controls, the onset of diabetes could thus differently affect methylation in cases and controls.

To assess the impact of aging, the baseline (50.9 ± 9.1 years) samples were compared to follow-up (same individuals 7.0 ± 0.9 years later**)**. A linear regression was then carried out using the time point (baseline vs. follow-up) as a predictor for methylation. The difference between the baseline and follow-up beta value was calculated as $\Delta beta = beta_{follow\text{-}up} - beta_{baseline}$ per marker and a linear model was fitted against Δbeta using the pairwise group status as a predictor (NGT vs. T-T2D, NGT vs. U-T2D, T-T2D vs. U-T2D).

In this analysis of age as a dichotomous trait, there were strong differences between baseline and follow-up within all three groups (**Figure 4.6).** Genomic inflation factors varied from λ=1.55 to λ=1.69.

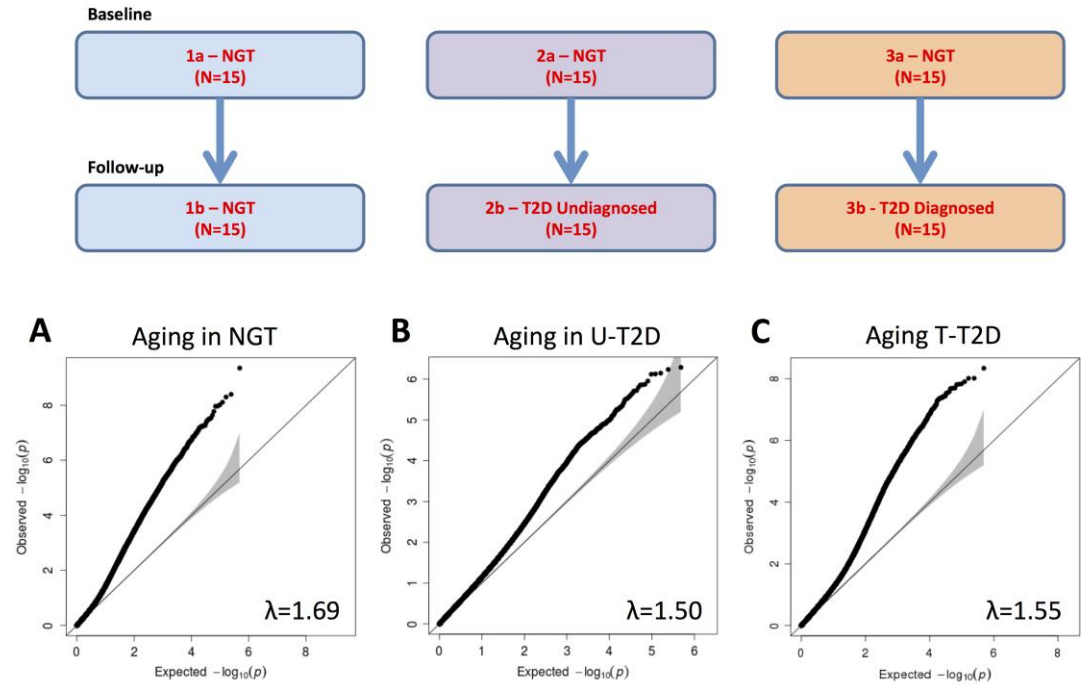


Figure 4.6: Association of epigenetic markers with age-related changes in the EpiMigrant Pilot Study. The QQ plots for p-values of linear models of methylation against time-point (baseline vs. follow-up) for all three groups are shown corresponding to the flowcharts above. Genomic inflation factors (λ) are given for all association tests are given in the right corner. A: NGT, B: U-T2D, C: T-T2D.

When calculating the change in methylation and comparing between groups, marked differences between people who developed T2D and those who remained NGT (**Figure 4.7**) were found. No differences between the treated and untreated T2D groups were found.

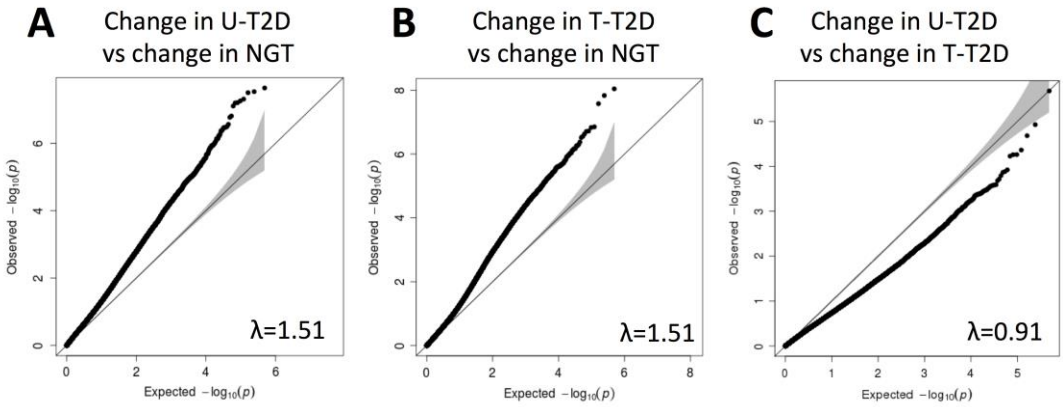


Figure 4.7: Comparison of changes in methylation between groups T2D in the EpiMigrant Pilot Study. The QQ plots for p-values of linear models of difference in methylation against group status for all three group combinations are shown. Genomic inflation factors (λ) for all association tests are given in the right corner. A: NGT vs. U-T2D B: NGT vs. T-2D, and C: U-T2D vs. T-2D.

The results suggest a systematic global effect of aging on methylation, in the NGT groups as well as T2D groups. This means that for the main study, prospective samples should be used, as the differential changes in methylation might be due to reverse causation. However, at this stage differences due to sample storage and preparation could not be ruled out.

4.2.3 Confounding factors and adjustment

Biological covariates can associate with DNA methylation and confound a study. It is thus important to assess whether the cell composition of a blood sample affects the measurement. For example, the mild inflammation and macrophage invasion of adipose tissue[268] associated with obesity can lead to spurious results due to cell differences. In blood, both T2D and obesity are associated with an altered WBC composition[286], which means that any epigenetic differences found may simply reflect

differences in cell types. Thus, I tested for association of WBC proportions as confounders in the Pilot Study.

Linear models were fitted to test for association of methylation with each of the WBC subset proportions (see Chapter 2.5.1 p. 40). Subsequently, the regression analyses of methylation against group status against were repeated including all the associated subsets as covariates.

Methylation at both time points also associated with white cell differentials and most strongly with lymphocytes (**Figure 4.8A**), neutrophils (**Figure 4.8B**) and eosinophils (**Figure 4.8C**). No association of methylation with monocytes (**Figure 4.8D**) or basophils (**Figure 4.8E**) was observed, potentially due to the low dynamic range for the subsets. Additionally, there was no association with total white blood cell count (**Figure 4.8F**). Associations of beta with T2D progression persisted after adjustment for these confounders (U-T2D vs. NGT: λ=1.46; T-T2D vs. NGT: λ=1.97).

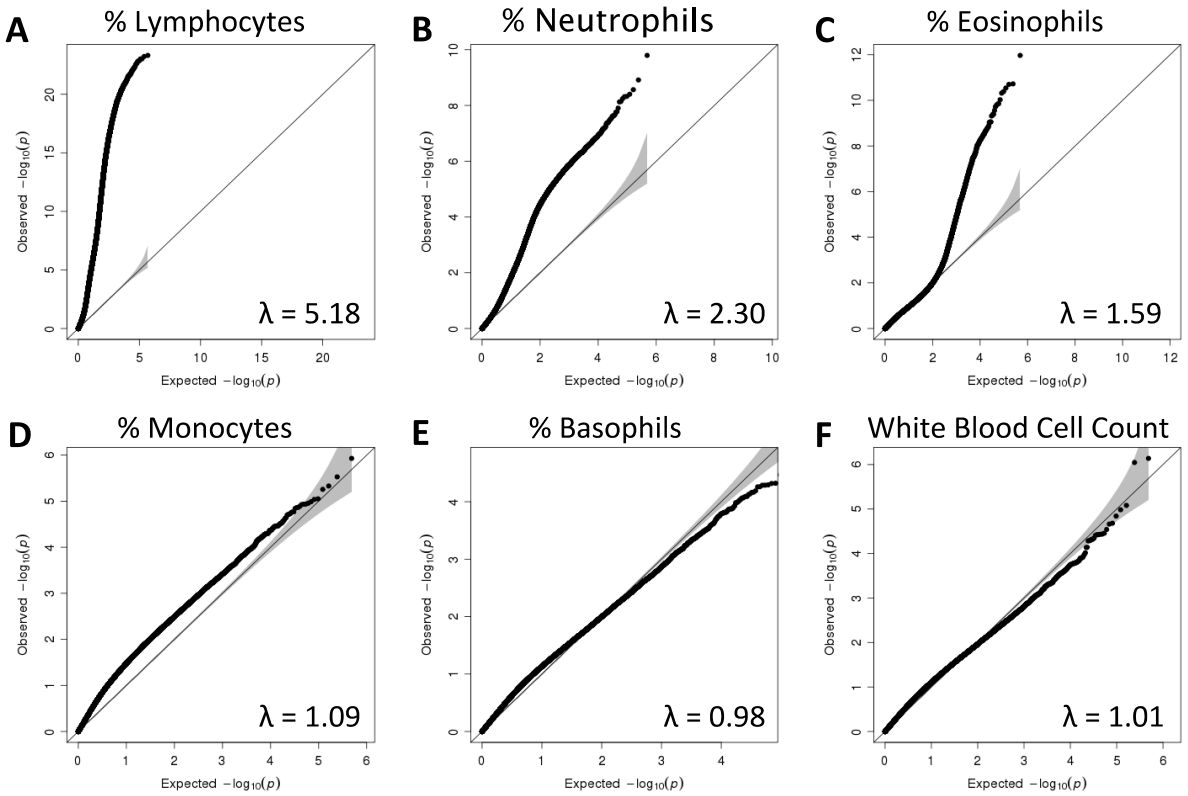


Figure 4.8: Association of epigenetic markers with white blood cell count and differential white blood cell subsets T2D in the EpiMigrant Pilot Study. The QQ plots for p-values linear models of methylation against white blood subsets for all 45 samples are shown. Genomic inflation factors (λ) for all association tests are given in the right corner. A: Percent lymphocytes, B: Percent neutrophils, C: Percent Eosinophils, D; Percent Monocytes, E: Percent Basophils, F: total white blood cell count.

WBC subsets are a strong determinant of methylation score, which is in line with previously reported cell-specific methylation profiles.[270] Differences in cell composition, caused for example by inflammation, could be a consequence of T2D and lead to reverse causation effects. Here, I found that WBC proportions strongly associate with methylation, which led to the decision to include only samples with WBC information into the EpiMigrant main study.

4.2.4 Pilot Study Conclusion:

Previous work has shown that methylation levels varies with time[179, 221] and between cell types[270], which might confound EWAS analysis if not corrected for. My work for the Pilot Study confirmed these observations with methylation data measured with the 450k array. I found systematic differences between the follow-up and baseline samples that were different between T2D and controls, which means retrospective case-control comparisons may not identify markers that predict future T2D. The strong association with aging also hints towards a large number of secondary effects (reverse causation) in methylation measurements, which I investigated in more detail in the main study. Moreover, WBC subsets are a strong determinant of methylation score, which is in line with previously reported cell-specific methylation profiles.[270] Carrying out a first smaller-scale methylation experiment, as well as understanding these confounding factors was important for design of the main EpiMigrant study, which generated data of sufficient size to be used for subsequent method and pipeline development.

4.3 EpiMigrant Main Study

4.3.1 Quality Control Parameter Thresholds and Metrics

As a first step in the analysis, I carried out quality control (QC) to assess the data quality. However, there is no consensus on which metrics and thresholds to use for 450k data. I thus had to devise methods to determine appropriate measures of

quality. Detection p-values, as developed by Illumina[280], are routinely used to filter data points, samples and markers (**Figure 4.9**). However, the thresholds set by Illumina are arbitrary and metrics needed to be found to determine which thresholds give the best payoff between removing data points and good data quality. Three rationales were used: 1) Females will not have true signals on chromosome Y, whereas male signals can be assumed to be true. An exclusion method maximising the male signals while minimising female signals would be most favourable. 2) Outlier signals are often generated from spurious measurements. If all markers captured by the QC are spurious, the outlier rate should be independent of the call rate (number of data points per marker). In the extreme case, markers with low call rate would have a high outlier rate as not all spurious signals would be caught. 3) Removing spurious signals would decrease noise in the data, and thus improve correlation between the 36 duplicate samples.

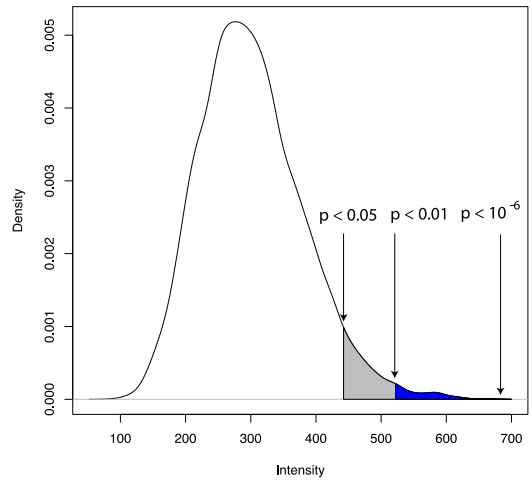


Figure 4.9: Distribution of intensities used for calculation of for detection p-values. The background probe distribution was randomly from sampled negative background control probes from the main study and three thresholds ($p<0.05$, $p<0.01$ and $p<10^{-6}$) are indicated illustratively. The individual detection p-values are calculated by comparing the measured signals at each probe to the background distribution.

I tested all these three methods for two detection p-values, the recommended $p < 0.05$ threshold and $p < 10^{-16}$, the minimum possible number in the analysis software to find an optimal QC threshold.[215]

The three metrics are implemented as follows:

1) The call rates on chromosome Y were calculated in males and females separately, and the percentage of total signals is compared between males and females.
2) To define a measurement as an outlier, the interquartile range (IQR) per marker was calculated. Beta values were then considered as outliers if they fall below quartile 1 – 1.5 × IQR or above quartile 3 + 1.5 × IQR. Outlier rates are calculated as the number of outlying beta values divided by the number of non-missing beta values.
3) The Pearson correlation between the 36 duplicates was calculated at each of the two thresholds.

Largely consistent with the expected absence of chromosome Y signals in females, 70% of methylation signals in females have a detection p-value above 0.05, while the same applies to only 0.2% of methylation signals in males (**Figure 4.10A**). These numbers increase further when the detection p-value threshold in decreased to $p<10^{-16}$ (88% in females, 0.28% in males). At this stringent detection p-value threshold, 27 markers have call rates above 0.9. The majority of these markers were identified to cross-hybridise with more than one region of the bisulphite-converted genome (**Figure 4.10B**).[220]

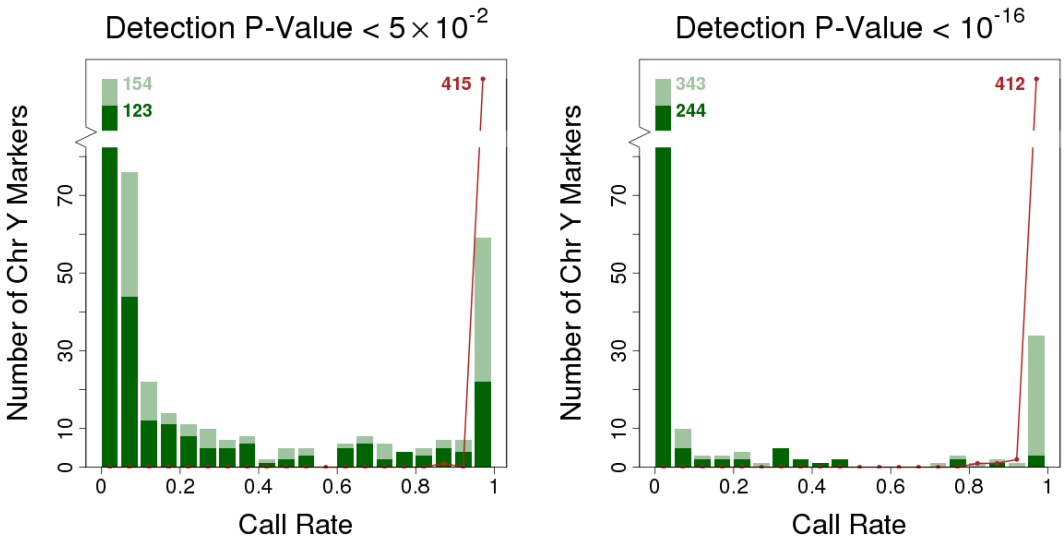


Figure 4.10: Distributions of Y-chromosomal call rates in females and males at two different detection p-value thresholds (0.05 and 10^{-16}). All measurements are taken exclusively from chromosome Y. Call rates were calculated as the proportion of non-missing values per marker at each detection p-value. The distribution of female call-rates is shown as a green histogram and the male distribution as a red line corresponding to the histogram bins. Light green bars correspond to counts prior to the removal of cross-hybridising probes. A: Detection p-value < 0.05 B: Detection p-value < 10^{-16}.

Considering all autosomal markers, more than 91% of methylation signals that were assigned a detection p-value above 0.05 are outliers by the above definition. However, markers with low call rates tend to have a higher number of outliers (**Figure 4.11**). At a detection p-value threshold of 0.05, markers with a call rate > 98% have an average outlier rate of 1.46%, whereas markers with a call rate < 98% have a significantly higher average outlier rate of 3.44% (T-test p-value = 2.2 x 10^{-16}). For more stringent detection p-value thresholds this difference becomes substantially smaller: at the maximum detection p-value threshold 10^{-16} the proportion of outliers is 1.55% (T-test p-value = 0.026).

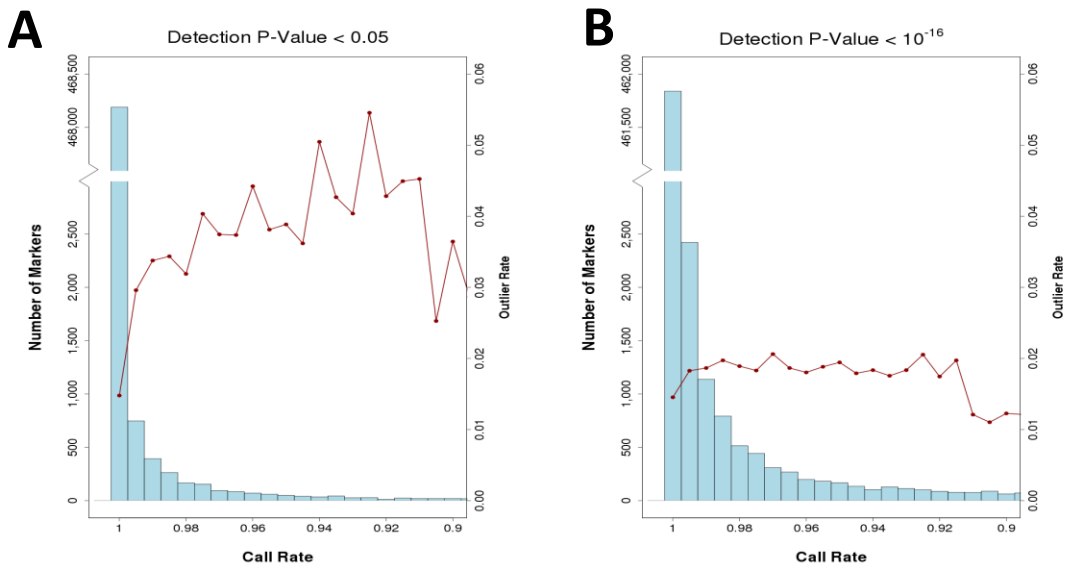


Figure 4.11: Distributions of autosomal call rate) and outlier rate (red line) at two different detection p- value thresholds (0.05 and 10^{-16}). All measurements are taken exclusively from chromosome 1-22. Call rates were calculated as the proportion of non-missing values per marker at each detection p-value. The distribution of autosomal call-rates is shown as a blue histogram and the outlier rate distribution as a red line corresponding to the histogram bins. A: Detection p-value < 0.05 B: Detection p-value < 10^{-16}.

When applying a detection p-value threshold of 10^{-16}, a significantly higher correlation between the 36 duplicate sample-pairs before and after data normalisation (paired T-test p-value 2.91×10^{-23}) is observed (**Figure 4.12**).

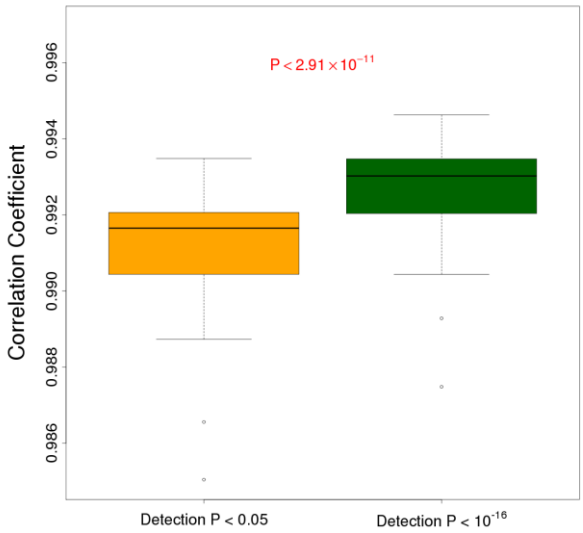


Figure 4.12: Sample-wise correlation across all CpG **markers at the two different detection p-value thresholds (0.05 and 10^{-16}).** The boxplots show the distribution of

pairwise Pearson correlation coefficients between the 36 samples measured in duplicate. Orange bar: p-value < 0.05 Green bar: Detection p-value < 10^{-16}.

On the basis of the above metrics, I chose to use a detection p-value threshold of 10^{-16} for all subsequent analyses. Sample call rates were recalculated which resulted in the exclusion of one further sample and a total number of 2,664 samples in the main study. The overall call rate still remains high, as the missingness rises from the initial ~0.2% to ~0.4% (**Figure 4.13**), further suggesting that the data is of high quality. I also found that cross-hybridising probes drive spurious signals on chromosome Y markers in females, which will be discussed further in section 4.3.3.

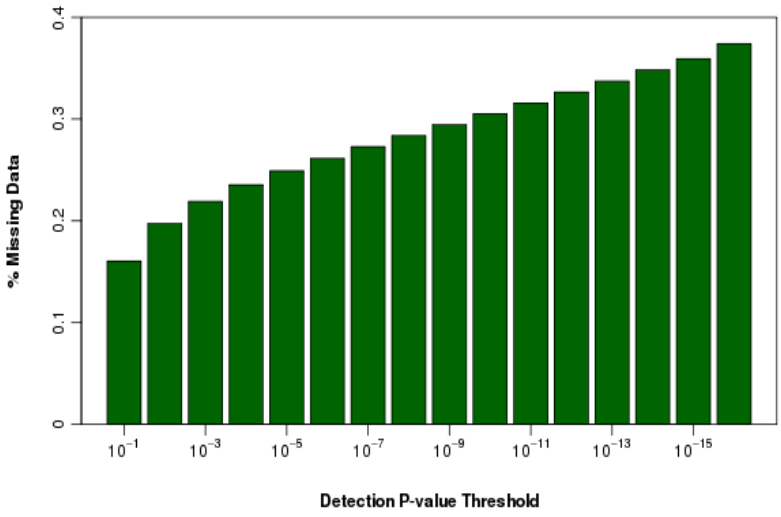


Figure 4.13: Percentage of missing data points at different detection p-value thresholds (0.1 to 10^{-16}). Measurements across all chromosomes were taken into account. Call rates were calculated as the proportion of non-missing values per marker at each detection p-value. The histogram shows the moderate effect of exponentially decreasing the p-value on the overall marker call rate.

4.3.2 Normalisation Methods

After excluding spurious signals, I aimed to improve data quality by normalisation of the methylation signals between arrays. This is a procedure that is routinely done to ensure that all measurements are on the same scale, and to remove variations due to global experimental factors. However, many different normalisation methods are available, which I first evaluated using the duplicate data set of 36 samples.

To assess the assay performance of each of the seven normalisation techniques (Chapter 4.3.2 p. 95), the Pearson correlation coefficient between the two sets of the 36 duplicate samples was calculated. As the duplicates arise from the same DNA sample after extraction, any variation is due to technical factors and higher correlation coefficients mean improved assay quality.

Before data normalisation markers and samples that failed QC were excluded and Illumina background correction was performed for all intensities values. A paired Wilcoxon-test was used to assess the difference between the normalisation methods. All normalisation methods were implemented using the R packages supplied with the publications.

The analysis revealed that QN(I) increases the correlation between repeated samples more than other published methods (**Figure 4.14**), followed by SWAN[216] and QN(B). The PBC[217] and BMIQ[218] methods perform similarly, as both methods focus on peak correction rather than between-array normalisation. A second method to assess the reproducibility is by calculating the Pearson correlation at each marker, rather than per sample.

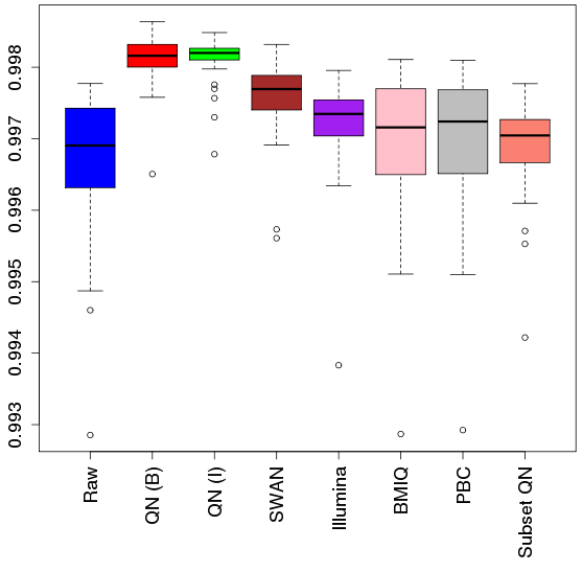


Figure 4.14: Pairwise duplicate correlation coefficients at each sample across all markers. The boxplots show the distribution of pairwise Pearson correlation coefficients between the 36 samples measured in duplicate. Raw: Non-normalised beta values, QN (B): quantile-normalisation of beta values, QN (I): quantile-normalisation of intensities in six categories, SWAN: Subset within-array normalisation, Illumina: Illumina Control Probe normalisation, PBC: Peak-based correction, BMIQ: Beta mixture quantile normalisation, Subset QN: Subset quantile normalisation.

Similarly, this analysis of the probe-wise reproducibility revealed that a large proportion of markers showed low correlation between the two duplicates in raw beta values (**Figure 4.15A**). After QN(I), the peak at r=0.8 increases. This means that the fraction of markers with high reproducibility increased. The remaining high proportion of markers with low correlation also has low inter-individual variance in this dataset (**Figure 4.15B**), and can thus not produce reliable measurements. Compared to non-normalised data in this analysis, as they mainly aim to remove probe-type bias.[217]

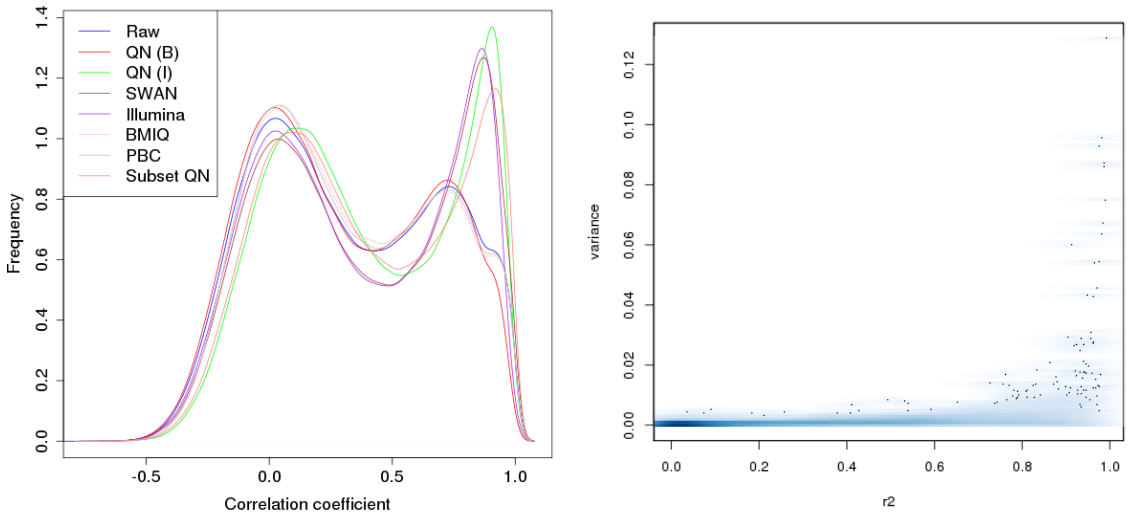


Figure 4.15: Marker-wise duplicate correlation coefficients at each marker across all pairs. A: The density plots show the distribution of pairwise Pearson correlation coefficients between the 36 samples measured in duplicate. Raw: Non-normalised beta values, QN (B): quantile-normalisation of beta values, QN (I): quantile-normalisation of intensities in six categories, SWAN: Subset within-array normalisation, Illumina: Illumina Control Probe normalisation, PBC: Peak-based correction, BMIQ: Beta mixture quantile normalisation, Subset QN: Subset quantile normalization B: Variance plotted against r^2 between replicated samples for all autosomal markers. The low correlation mode in graph A is caused by low variance and thus noise at a large number of markers.

I found that out of the tested methods for normalisation QN(I) performs best both in the per-marker and per-sample analysis of reproducibility. SWAN[216] and QN(B) performed significantly worse. As QN(I) increased both the sample-wise and the probe-wise correlation, further analyses were based on the quantile-normalised dataset.

4.3.3 Probe QC Categories

As described earlier, the design of the 450k array includes markers, which may have inherent technical issues. With the development of the workflow in mind, and after deciding on a normalisation method, the next question was whether inclusion of potentially problematic markers has an impact on the normalisation by introducing noise. Five different types of probes were considered:

1) Cross-hybridising probes: the 50-mer probes might hybridise to another similar sequence in the genome (cross-hybridise) and measure signals from more than one genomic location.
2) Probes with SNPs: probe sequences may have SNPs underlying the target sequence, which can potentially give rise to spurious differential signals due to different genotypes. The SNPs need to be relevant to the South Asian population investigated.
3) Probes with low call rate (Chapter 4.3.1, p. 90): it was postulated that probes leading to low call rates across all samples might be inherently problematic
4) Probes with low outlier rate (Chapter 4.3.1, p. 90): a low outlier rate may also reflect an overall poorly performing probe.
5) Probes measuring methylation at a CpA site: these probes measure methylation at the CpA dinucleotide, for which there is evidence of differential methylation in human embryonic stem cells.[287, 288]
6) Probes lying within the MHC region: the Major Histocompatibility Complex (MHC) region on chromosome 6 is a region of high variability with several different haplotypes.[289] It is therefore possible that probes targeting this region might give rise to results confounded by the complex haplotype structure.

For each of the above categories, probes were annotated as following:

1) A probe was considered cross-hybridising if its probe sequence mapped to more than one genomic location (N=39,963, previously identified by Price et al.[220])
2) Probes on the 450K array that have at least one SNP located within the probe sequence with MAF > 1% based on genotyping data of 1,909 Indian Asians (71% of individuals in the primary study, generated by Dr Benjamin Lehne) were annotated as SNP probes.

3) As described in Chapter 4.3.1 (p. 90) call rates based on detection p-values were determined for each methylation marker. Any markers with a call rate below 98% were flagged as low call rate markers.
4) Outlier rates (Chapter 4.3.1 p. 35) for each marker were determined using the 1.5 x IQR definition. Any markers with an outlier rate above 5% were flagged as high outlier rate markers.
5) The 450k array contains a number of probes that measure methylation at CpA rather than CpG sites (N=2,995), which are flagged in the annotation.
6) The MHC region is located at chromosome 6 position 29.62mb – 33.16mb (hg19).[289] Any probes that are annotated to map within this position are considered MHC probes (N=9,527).

The total amount of probes in each category is given in **Table 4.2**. To investigate the effects of markers in different categories on normalisation, QN(I) was performed including all markers and separately within each marker category. As described above, the accuracy of the normalisation is assessed by the per-sample and per-marker correlation between duplicate pairs.

Table 4.2: Number of markers on the 450k array in different categories.

Marker category	Number of markers
Chromosome X Markers	11,135
Chromosome Y Markers	416
Autosomal Markers	473,961
Total QC Markers	105,766
Markers with SNPs	56,661
Cross-hybridising Markers	39,963
Low Call Rate Markers	4,684
High Outlier Rate Markers	5,383
Markers in the MHC region	9,527
CpA Markers	2,994

When considering markers that are not part of any exclusion category, a significantly lower per-marker correlation for QN(I) including all markers is observed (**Figure 4.16**). In contrast, there is a significantly higher per-sample correlation for QN(I) including all markers. Across all autosomal markers an improvement in the per-marker correlation between duplicates when QN is performed including all markers and no significant difference in the per-sample correlation is achieved.

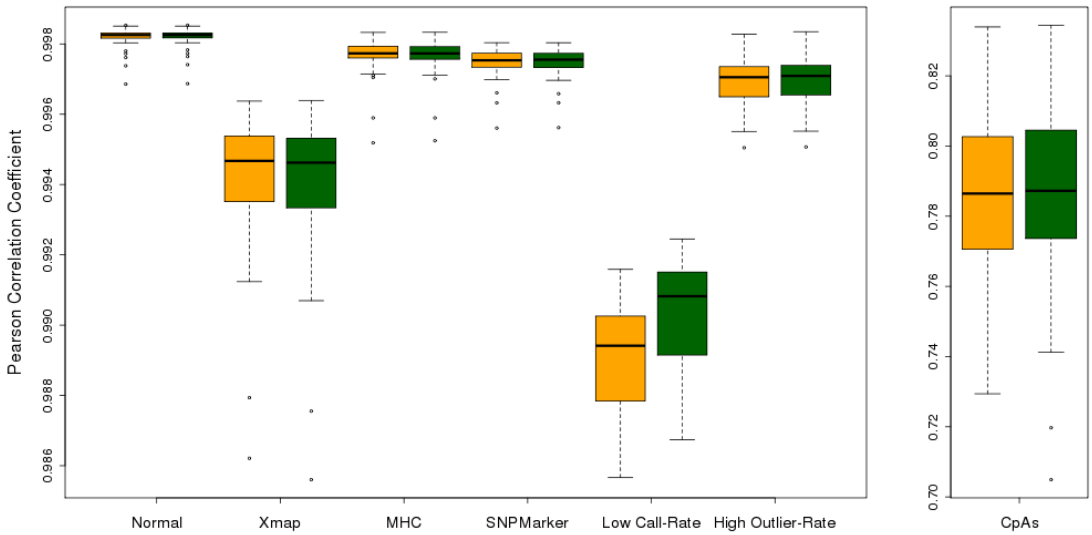


Figure 4.16: Correlation between duplicates for markers in each category. The boxplots show the distribution of pairwise Pearson correlation coefficients between the 36 samples measured in duplicate before and after QN(I). Orange boxes: before QN(I), Green boxes: after QN(I), Raw: Non-normalised beta values, Orange boxes: before QN (I) .Normal: Autosomal markers excluding other QC criteria. Xmap: Cross mapping markers. MHC: Markers in the major histocompatibility region. SNPMarker: Markers of which the probe target sequence contains a SNP. Low Call Rate: markers with a call rate <98%. High Outlier Rate: markers with an outlier rate 5%

The inclusion of markers within the QC categories did not seem to negatively impact the normalisation. Subsequent analyses are therefore based on QN(I) of all autosomal markers. The discovered marker categories are then used to annotate results and flag potentially problematic markers.

4.3.4 PCA and outlier detection

If the samples cluster by biological phenotypes, such as case-control status, global effects could confound single marker analyses. Here, I performed a principal

component analysis (PCA) to capture these global effects and also to identify outliers.

A PCA was carried out using 450k methylation array data for all samples (N=2,664) in the study and all markers that had no missing data (>60%) using the subset of filtered samples and autosomal markers. The first five principal components (PCs) were compared pairwise to detect clustering with gender (**Figure 4.17A**) or prospective T2D status (**Figure 4.17B**)

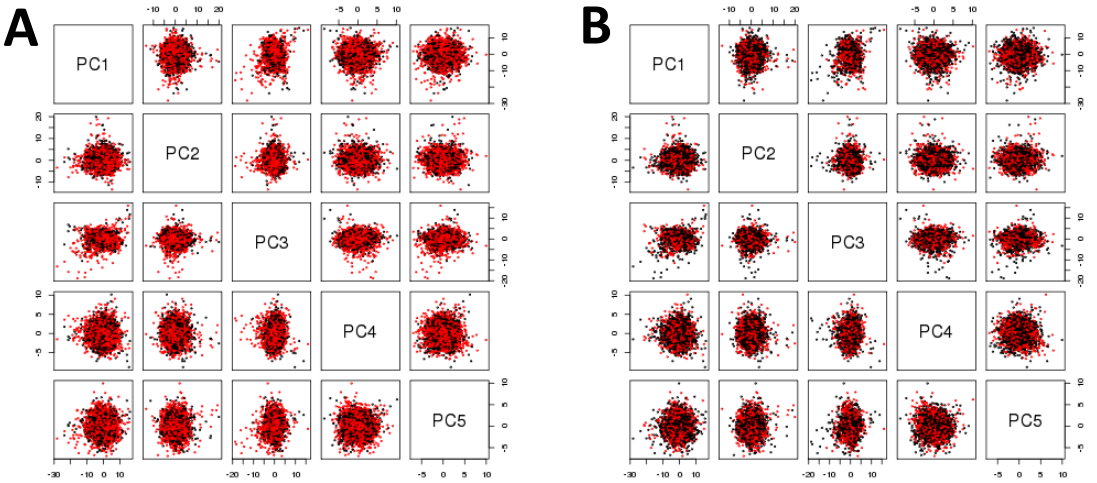


Figure 4.17: Principal Component Analysis. PCs where calculated including markers with no non-missing values as described in Chapter xx p. xx. Scatterplots depict all combinations of PC1 to PC5. A: Clustering by gender, red dots correspond to males and black dots to females. B: Clustering by incident T2D status, red dots correspond to T2D cases and black dots to controls.

No apparent clustering was observed between the first five principal components and gender or future T2D status. While PC1 explains 0.10% of the variance, the variance explained by each component drops below 0.01% beyond the fifth component (**Figure 4.18**). Based on the first two PCs, no outliers were identified and all 2,664 samples were retained.

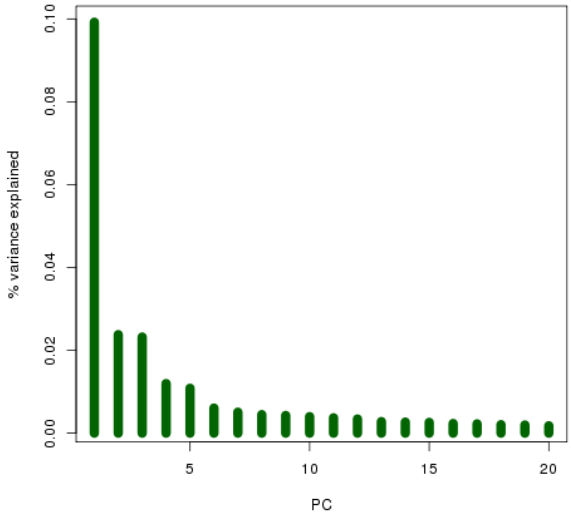


Figure 4.18: Variance Explained by each PC. Histogram showing the percentages of variance explained by each of the first 20 PCs. While the first PC explains ~10% of the variance in the data, the variance explained drops after PC5.

Age and gender do not seem to influence global methylation patterns. No samples were excluded based on this PCA.

4.3.5 Global Correlation Structure and the Null Hypothesis

It has previously been shown that methylation patterns can correlate over long ranges[163], which can lead to lead to statistical problems due to multiple testing (both false positives and negatives). A null hypothesis can be established by permutation of the data points, and correlation patterns across the epigenome can be investigated by randomly sampling markers.

Case-control labels were permuted amongst samples and a logistic regression was carried out for each marker. This process was repeated 1,000 times giving rise to 1,000 sets of p-values under no association. The permuted data sets were generated by Dr Benjamin Lehne (Imperial College). Secondly, to further quantify the correlation, 1,000 autosomal markers were randomly chosen and the Pearson correlation between beta values of all possible marker pairs were calculated.

Under no association, a statistical deflation is observed for the majority of permutation p-values (λ_{median}=0.91; $\lambda_{2.5\%tile}$=0.57) whilst a small number of permutations show a high degree of statistical inflation ($\lambda_{97.5\%tile}$=2.91, (**Figure 4.19A**). When beta values for each marker are randomly re-assigned amongst samples, no bias towards deflated test-statistics (λ_{median}=1.00) and a symmetrical distribution of p-values around the expected ($\lambda_{2.5\%tile}$=0.99; $\lambda_{97.5\%tile}$=1.01, (**Figure 4.19B**) is observed. These results can be explained by a high degree of correlation between methylation markers, which reduces the number of independent tests. The theoretically expected p-values are based on the assumption of independence for each test. As a consequence, comparing expected and observed results in an apparent deflation of p-values. Correlation also explains the strong inflation that is observed for a small number of permutations: by chance a given marker might appear to be differentially methylated between randomly assigned cases and controls. Correlation between that marker and many other markers will result in an apparent statistical inflation. Randomly re-assigning beta-value amongst samples (for each marker) removes the correlation between markers.

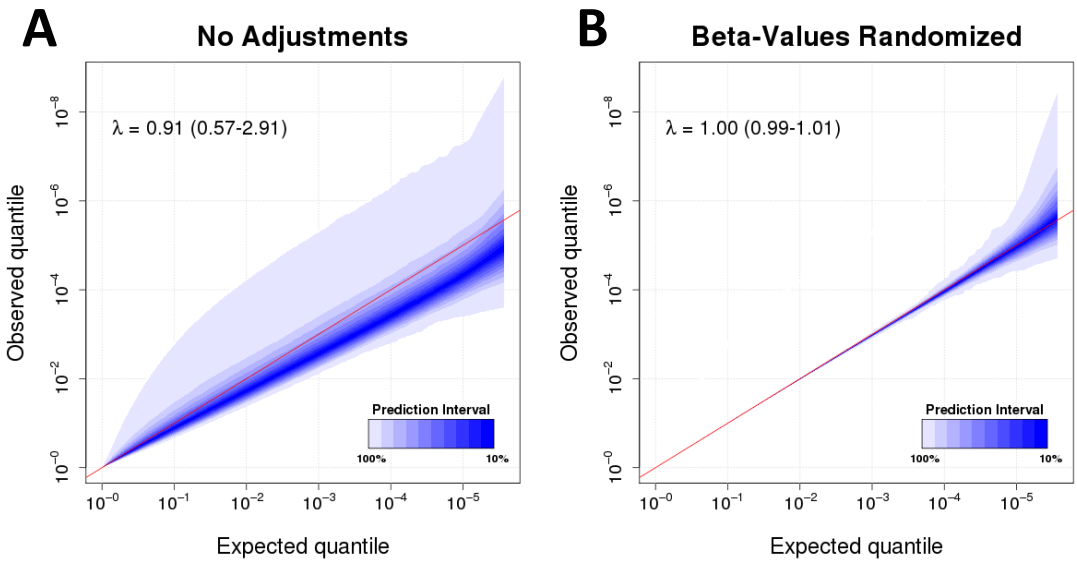


Figure 4.19: Distribution of p-values under the null hypothesis. The QQ plots allow for comparison of the theoretically expected p-value distribution under no association (randomly sampled variable) with the 1,000 permuted p-value distributions in a QQ plot whereby percentiles closer to the median are coloured in

darker shades of blue. A: Distribution of randomly permuted p-values based on 1,000 permutations of sample labels. B: Distribution of randomly permuted p-values based on 1,000 permutations of sample labels carried out at each marker separately.

This issue is further illustrated by the high degree of correlation between randomly chosen markers from all autosomes, which have a median Pearson correlation coefficient of r_{median}~ 0.2 (**Figure 4.20A**) and a standard deviation of SD~0.24. In contrast, when beta-vales are randomly re-assigned amongst samples, the correlation coefficients are distributed very closely around 0 (r_{median}~0; SD=0.017; **Figure 4.20B**).

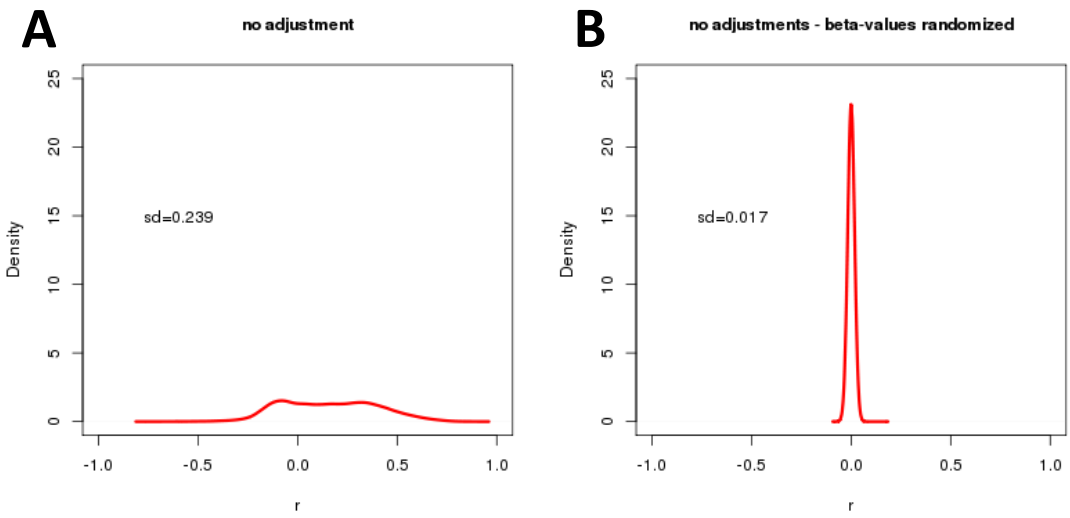


Figure 4.20: Correlation between randomly chosen CpG **markers.** To test if methylation correlates across the epigenome, 1,000 markers were randomly chosen and the Pearson correlation coefficients between all marker pairs were calculated for both the permuted and unpermuted data set of 2,664 samples. The resulting distribution of Pearson correlation coefficients is shown in red. A: unpermuted data set B: marker-wise permuted data set.

The results indicate that the broad prediction interval and departure from the theoretically expected p-values under the null hypothesis is caused by the high correlation of methylation across the genome. The next aim of this work was thus to recover the null hypothesis and maximise the independent number of tests by identifying factors driving the correlation structure.

4.3.6 Factors driving the correlation structure and batch effects

Confounding factors, such as technical artefacts that affect a large number of markers, can influence the null hypothesis by creating correlation between individual markers. In this section, I will discuss three ways to capture technical variation. The effect of the adjustment for technical factors will then be discussed in section 4.3.7.

1) Metrics from the control probes on the Illumina 450k array: The 450k array assesses multiple technical metrics of the experiment with a series of control probes. For example, a set of control probes measures methylation at sites that are known to be invariably methylated, thereby assessing the effectiveness of the bisulphite conversion.[191]
2) White Blood Cell Proportions: DNA methylation was measured in whole blood, which represents a mixture of different cell types. Due to the cell-specific nature of DNA methylation, differences in the cell composition can result in differences in beta values, as observed in the Pilot Study. In addition to the measured subsets a complementary set can be estimated based on the methylation data itself using an approach published by Houseman et al.[270]
3) Principal Components: I investigated whether any of the identified factors driving global correlation, or others, are reflected in the principal components of the global methylation data.

Control Probes: To adjust for technical biases, the signal intensities for 24 different types of control probes (grouped into 8 categories; **Table 4.3**) were retrieved. A set of PCs was then generated by Dr Benjamin Lehne from the control probe intensities (Chapter 2.1.6, p. 38), as these capture most of the experimental variation while keeping the number of variables included in the model at a minimum.

Table 4.3 Number of control probes on the 450k array in different categories

Category	Number of probes
Bisulphite conversion	10
Staining	2
Extension	4
Hybridisation	3
Target removal	2
Specificity	9
Non-polymorphic	4
Normalisation	186

Bisulphite conversion: control probes measuring efficiency of bisulphite conversion of genomic DNA, Staining: control probes measuring efficiency of staining in red and green colour channels, Extension: control probes measuring the efficiency of extension of A,T,C,G nucleotides, Hybridisation: control probes measuring the efficiency of hybridisation of synthetic targets. Target removal: control probes measuring the efficiency of the stripping step after the extension. Specificity: control probes measuring non-specific primer extension, Non-polymorphic: control probes measuring amplification of non-polymorphic regions of the genome, Normalisation: control probes measuring methylation in housekeeping genes.

To quantify the batch effects captured with the control probes, methylation values between the first and the second measurement (replicate 1 and replicate 2) for the 36 duplicate samples were compared. For each marker, a paired linear regression of methylation against the replicate status was performed (see Chapter 2). To adjust for the experimental effects, an increasing amount of control probe principal components were included in the model and statistical inflation was measured.

WBC Subsets: To adjust for potential confounding through white blood cell compositions, the total number of white blood cells and the proportion of five different subsets (lymphocytes, monocytes, neutrophils, eosinophils and basophils) were measured directly. In addition, complimentary CD8T, CD4T, Natural Killer, B-cells, monocyte and granulocyte proportions were estimated based on the Houseman et al.[270] approach. This method predicts proportions of cell type subpopulations based on methylation signatures from sorted human cells.[290] To assess the quality of the imputed subsets, the Pearson Correlation coefficients between the measured and estimated proportions were calculated. It was then tested whether inclusion of the

imputed white blood cell data in the models would have an impact on the correlation between 1,000 randomly sampled markers.

Principal Components: Lastly, it was investigated whether any of the identified factors driving global correlation are reflected in the principal components of the global methylation data. As described above, a PCA was carried out and the resulting first 10 PCs were investigated for correlation with confounding factors.

Control Probes: Despite the overall high correlation of beta values between duplicate pairs (r=0.998, see above) the regression reveals a high degree of statistical inflation (genomic inflation factor λ=2.11; **Figure 4.21**) Because samples were measured in duplicate these differences can only be attributed to technical biases.

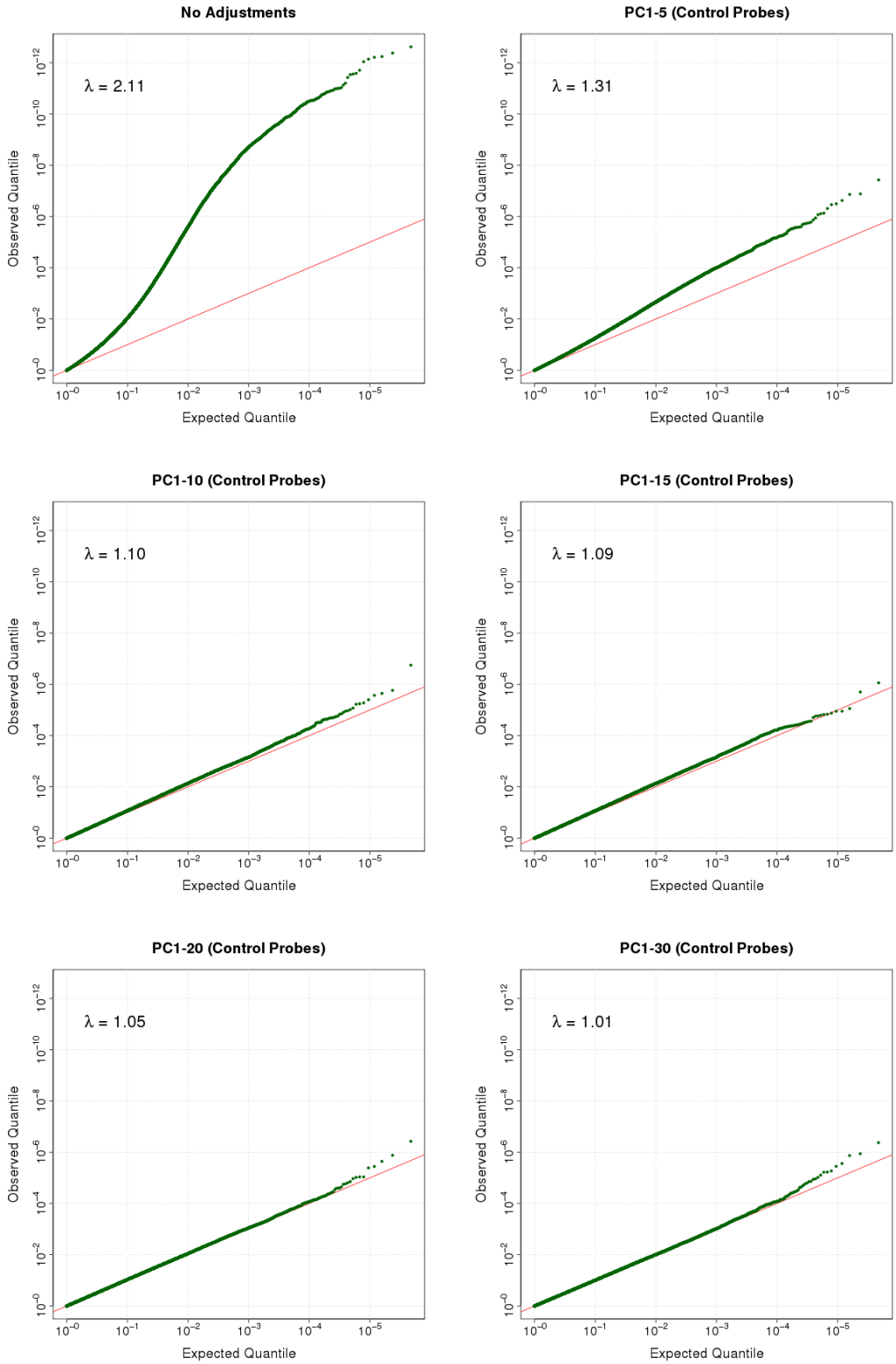


Figure 4.21: Regression of methylation against replicate status with control probe adjustments. The QQ plots for p-values linear models of methylation against replicate status (1 or 2) for the 36 duplicate samples are shown as green lines. An increasing number of control probe PCs were included as covariates. Genomic inflation factors (λ) are given for all association tests are given in the left corner.

Correction for 30 control probe predictors corrects for the majority of the inflation, resulting in a genomic inflation factor of $\lambda = 1.01$. The effect of control probe adjustment in the larger population study was assessed by carrying out a PCA on the QN(I) data analogously. Whereas before adjustment strong correlations with recorded technical factors, such as batch and plate were observed, control probe adjustment removes all these associations (**Figure 4.22**), showing that the control probe PCs can effectively account for batch effects.

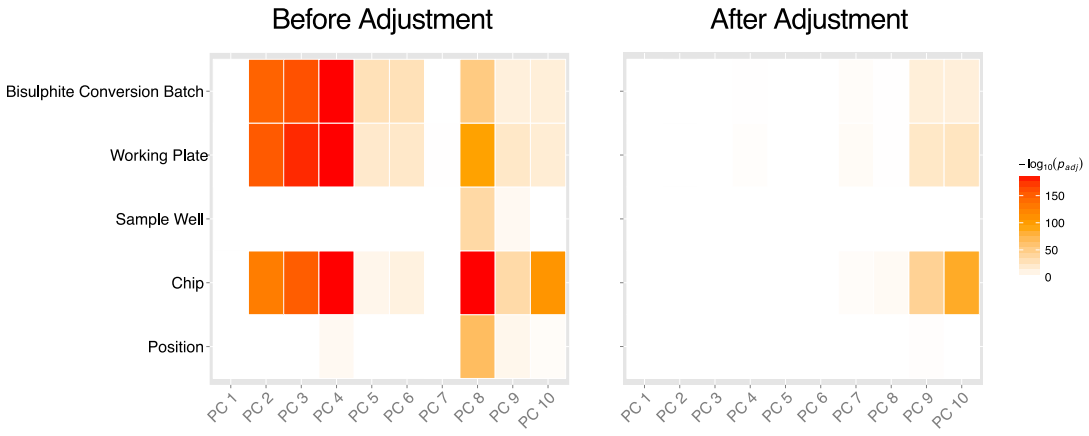


Figure 4.22: PCA before and after adjustment for control probes. The heat map shows the p-values from the pairwise regression of the first ten PCs against selected technical variables before (left) and after (right) adjustment for control probe PCs. Correction for control probes effectively removes technical biases.

WBC Subsets: Strong correlation is observed between measured and estimated WBC subsets for lymphocytes (measured lymphocytes compared to the sum of estimates for CD8T, CD4T, Natural Killer and B-cells; **Figure 4.23A**) and Granulocytes (the sum of measured neutrophils, eosinophils and basophils compared to estimates for granulocytes; **Figure 4.23B**). Monocytes were less correlated (measured monocytes compared to estimated monocytes; **Figure 4.23C**), which may be explained by measurement errors due to the lower overall proportion for this cell type.

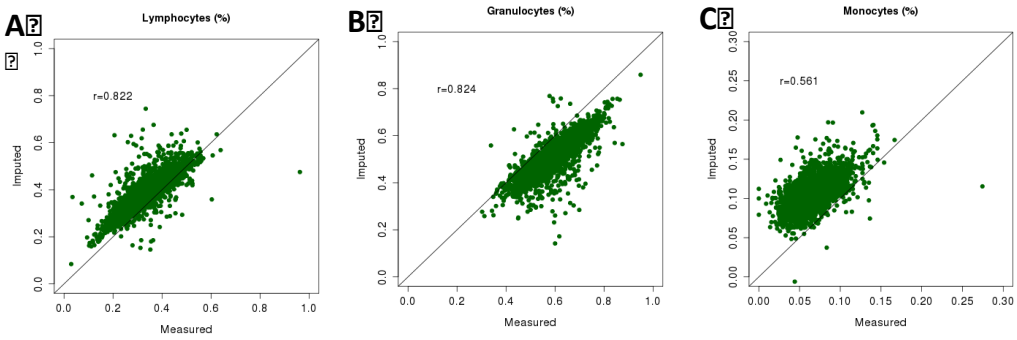


Figure 4.23: Correlation of measured with imputed white blood cell subset proportions for lymphocytes, granulocytes and monocytes. Scatterplots showing paired measurements for all 2,664 samples between imputed and measured WBC proportions. The Pearson correlation coefficient is given in each plot. A: Percentage Lymphocytes. Imputed lymphocytes are the sum of CD4+, CD8+, NK cell and B cell counts. B: Percentage Granulocytes. Measured granulocytes are sum of the basophil, eosinophil and neutrophils counts. C: Percentage Monocytes.

Adjustment for measured total WBC count significantly reduces the correlation between markers (**Figure 4.24**). With respect to WBC subsets a stronger decrease in correlation is observed when adjusting for estimated WBC subsets compared to measured WBC subsets (**Figure 4.24**).

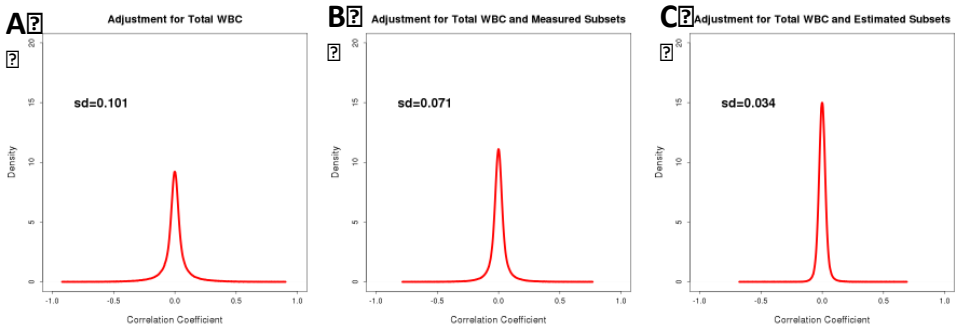


Figure 4.24: Correlation between 1,000 randomly chosen markers after adjustment for WBC parameters. 1,000 markers were randomly chosen and the Pearson correlation coefficients between all marker pairs were calculated for both the permuted and unpermuted data set of 2,664 samples. The resulting distribution of Pearson correlation coefficients is shown in red, corresponding to beta values after adjustment total WBC, total WBC and measured subsets, and total WBC and estimated subsets.

PCA: As shown in **Figure 4.25**, a number of confounding factors associate with the principal components. Most notably, the first three principal components highly correlate with both measured and estimated WBC subsets (**Figure 4.25**).

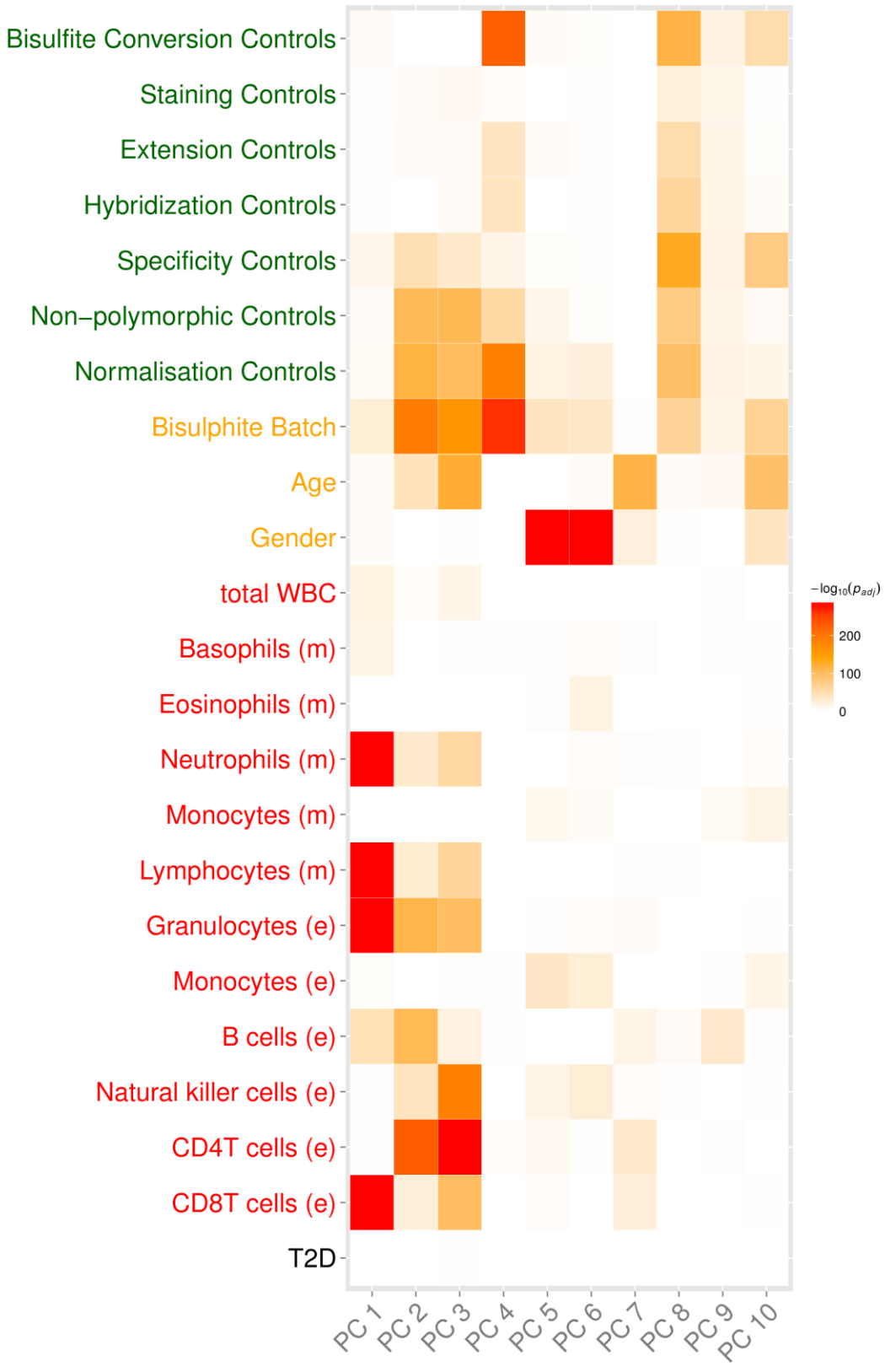


Figure 4.25: Associations between principal components and WBC parameters. The heat map shows the p-values from the pairwise regression of the first ten PCs against measured (m) and estimated (e) with WBC subsets. Colours represent P-values for correlation of different factors with PCs 1-10.

In summary, the analysis of the duplicate samples suggests that including up to 30 control probe PCs into the model decreases the batch effect, and it was decided to use up to 30 PCs in the following analyses. I will further investigate the effect of the adjustments on real signals in the following sections. Adjustment for estimated WBCs is more effective in reducing correlation than adjustment for measured WBCs. This might be explained by the fact that estimated WBC subsets report proportions for six different subsets, including four lymphocyte subclasses. In contrast WBC subsets were measured for five slightly different cell populations and lymphocyte subclasses were not measured separately. As a consequence of these results, estimated WBC subsets are included in the primary regression model. The observation that the first five PCs correlate with multiple confounding factors and the fact that the variance explained by each component drops below 0.01% beyond the fifth component (Figure 4.18) led to the decision to include the first five principal components as covariates in the main analysis.

4.3.7 Adjustment for Confounding Factors

After identification of the major confounding factors that drive correlation in methylation, the effect of adjustment for these factors on the null hypothesis needed to be addressed. As shown above, correlation can lead to statistical inflation, which is aggravated by the fact that the confounding factors affect multiple correlated markers at once.*:*

The permutation analysis was repeated by carrying out 1,000 sets of logistic regressions for each methylation marker, predicting the randomised case-control status as a function of the beta-value adjusted for set of linear predictors described above.

For raw beta values, a deflation of test statistics with a large prediction interval (**Figure 4.26A**) is observed, which is reduced by quantile normalisation (**Figure 4.26B**), suggesting that quantile normalisation also corrects for technical biases and batch effects. Adjustment for control probes and the bisulphite batch variable has only a small effect on statistical inflation (**Figure 4.26C**), but reduces correlation, thereby increasing the effective number of statistical tests. Including gender and age, has has little effect on the permuted distribution of p-values (**Figure 4.26D**). The most notable change is apparent when the estimated WBC counts are included into the model, which bring the permuted distribution close to the theoretically expected distribution (**Figure 4.26E**). The prediction interval is further narrowed by inclusion of the first five PCs of the main data, meaning that a model closer to the null hypothesis is recovered (**Figure 4.26F**). As described previously, these observations can be explained by global correlation between methylation markers (**Figure 4.27**).

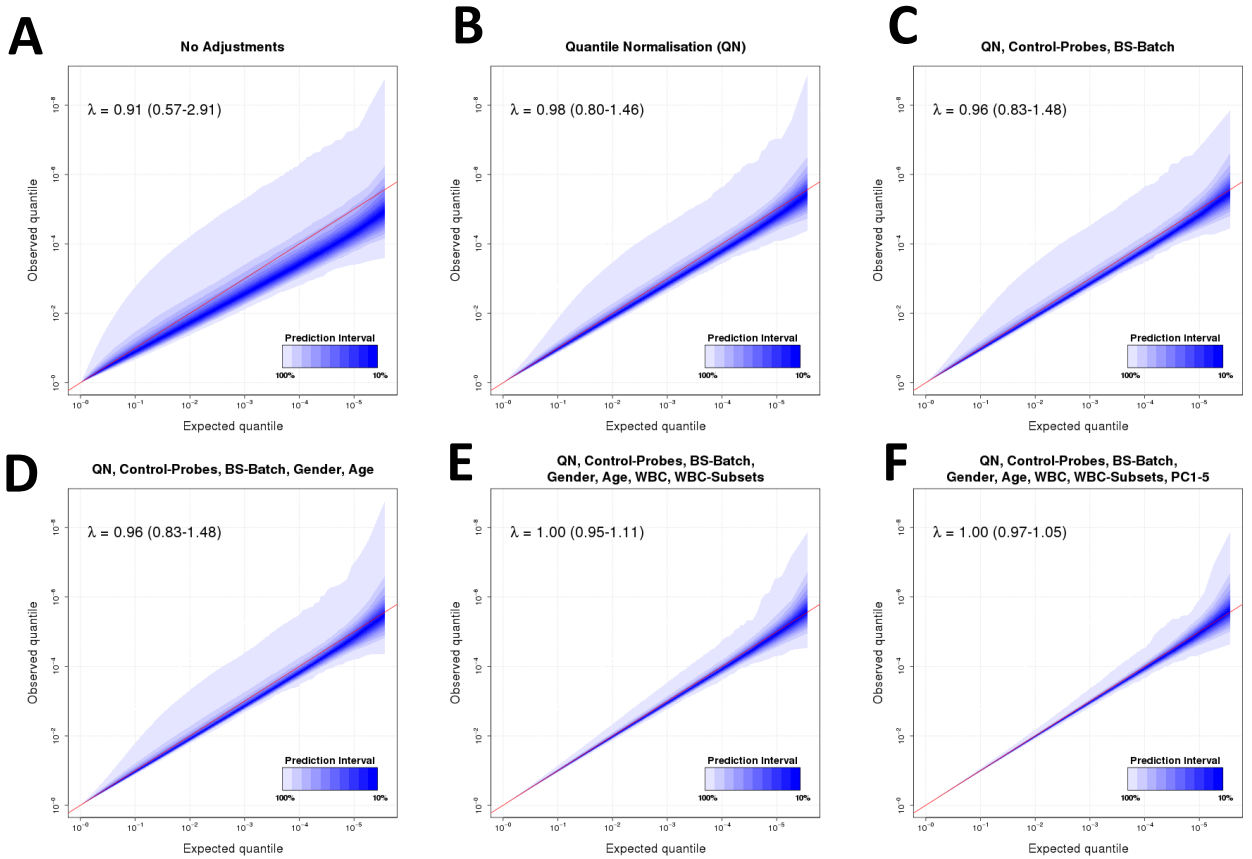


Figure 4.26: Permutation p-value distributions after different adjustments of beta values. The QQ plots allow for comparison of the theoretically expected p-value distribution with the 1,000 permuted p-value distributions in a QQ plot whereby percentiles closer to the median are coloured in darker shades of blue. A: No Adjustments, B: QN (I), C: QN(I) and bisulphite batch (BS), D: QN(I) ,bisulphite batch (BS), gender and age, E: QN(I) ,bisulphite batch (BS), gender and age, total WBC and estimated WBC subsets, F: QN(I) ,bisulphite batch (BS), gender and age, total WBC, estimated WBC subsets and PC1-5.

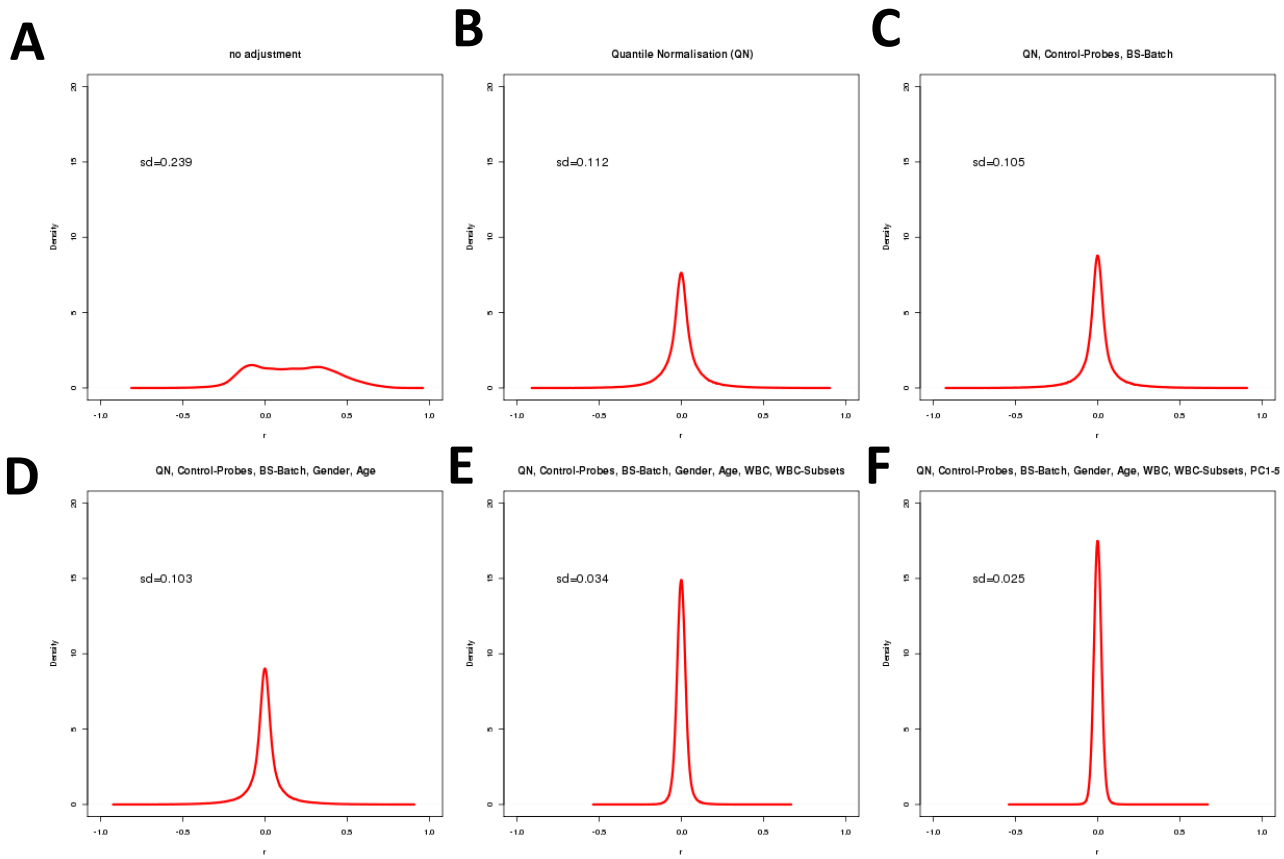


Figure 4.27: Correlation between randomly chosen markers after different adjustments of beta values. 1,000 markers were randomly chosen and the Pearson correlation between all marker pairs were calculated on the residuals after fitting a linear model including total WBC count, measured WBC subset proportion and estimated WBC subset proportion for all 2,664 samples. The resulting distribution of Pearson correlation coefficients is shown in red. A: Pearson correlation coefficients based on residuals after adjustment for total WBC, B: Pearson correlation coefficients based on residuals after adjustment for total WBC and measured WBC subsets, C; Pearson correlation coefficients based on residuals after adjustment for total WBC and estimated subsets.

Adjustment for WBCs reduces between-marker correlation, which increases the genomic inflation factor λ and results in a narrower prediction interval. In addition, since WBCs proportions can be affected by T2D or obesity, adjustment can reduce spurious associations, in particular for markers strongly associated with WBC counts. These results lead to the decision to adopt the adjustment for the linear predictors described above into the analysis step of the final pipeline.

4.3.8 Adjustment for Confounders and Local Correlation Patterns

Whilst adjustments and normalisation effectively reduce global correlation patterns, I also assessed if there is an effect on the local correlation between *cis*-markers, to ensure that non-spurious correlation patterns are preserved. Previously published data using the 27k array suggested that within short regions (>1.5 kb), there should be an enrichment of highly positively correlated markers.[179]

Local correlation was determined for all possible pairs of autosomal markers up to 5,000 bp apart. Distance between markers was based on the annotated position of the CpG sites on the forward strand. Pearson correlation coefficient between marker pairs were calculated based on beta values (raw) and residuals derived after adjustment for confounding factors. The correlation is represented graphically on a continuous scale using a sliding 300 bp mean.

Within the 450k data (see Chapter 2), there is a high "baseline" correlation of r=0.2 across larger distances, which only increases slightly as the distance between markers approaches 0 (**Figure 4.28A**). However, after QN(I), this global correlation is reduced to a distribution close around r=0, as expected, with a subtle increase as closer distances (**Figure 4.28B**). This means that the high correlation baseline across all markers is most likely due to technical confounding, which is confirmed by adjusting the raw data for control probes and bisulphite batch.

Average correlation at shorter distances could be diluted by markers with poor assay performance and low variance in the above analysis. To assess this, pairwise correlations at the 10,000 most variable markers on all chromosomes were investigated. In this analysis, the previously published results[179] are replicated with the raw (**Figure 4.28C**) data. This again highlights the low inter-individual variance in methylation based on whole blood samples. Consistent with the previous observations, QN(I) and adjustments for linear predictors in turn lower the long-range correlation, but preserve local correlation patterns (**Figure 4.28D**).

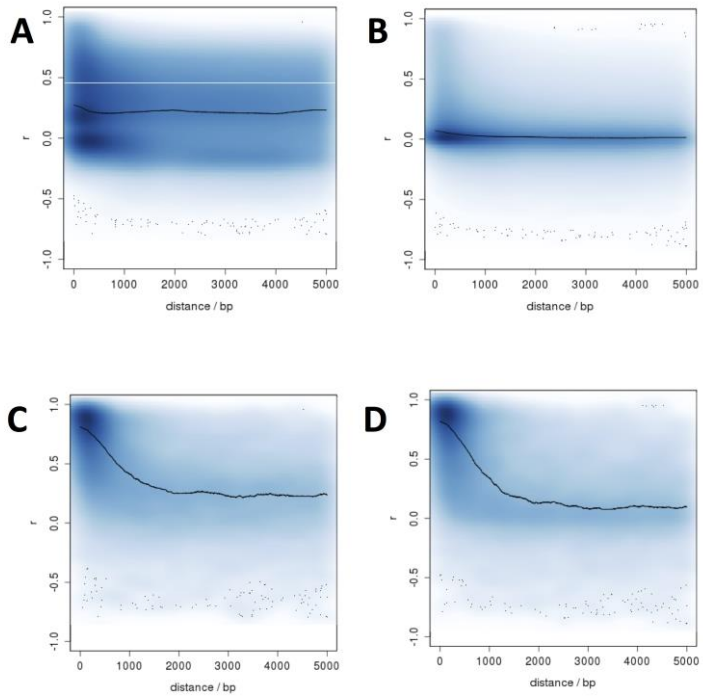


Figure 4.28: Local Correlation between CpG sites up to 5000 bp apart. The Pearson correlation coefficients for all possible CpG marker pairs up to 5000 bp were calculated and are shown as smooth scatter plots representing the density of points at any distance. The analysis was repeated selecting only the 10,000 most variable markers. A: Pairwise correlation between all markers against distance before QN (I), B: Pairwise correlation between all markers against distance after QN (I), C: Pairwise correlation between the 10,000 most variable markers against distance before QN (I), D: Pairwise correlation between the 10,000 most variable markers against distance after QN (I).

My results show that adjustments for linear predictors lower the long-range correlation, but preserve local correlation patterns, consistent with reported

correlations within 2 kb reported by previous studies.[179] This provided further evidence that QN(I) and adjustment for confounders does retain informative signals while removing global correlations.

4.4 Assembly of the Final Pipeline

All of my results above were integrated into a final pipeline in collaboration with Dr Benjamin Lehne using a combination of shell scripts and R for use on a PBS (Portable Batch System) cluster environment (**Figure 4.29**). Signal intensities are filtered for multiple QC criteria followed by QN(I) and calculation of beta values. In addition, intensities from Illumina control probes are used to derive PCs (control probe PCs). Based on beta values, proportions of WBC subpopulations are estimated and PCs are derived from intermediary residuals. Beta values are corrected for WBC proportions, control probe PCs, study-specific variables and PCs. The residuals derived from this model are corrected for technical and biological biases and can be used to test association of DNA methylation with a phenotype. This approach allows for the analysis of DNA methylation data and its integration with other datasets in a computationally highly efficient manner.

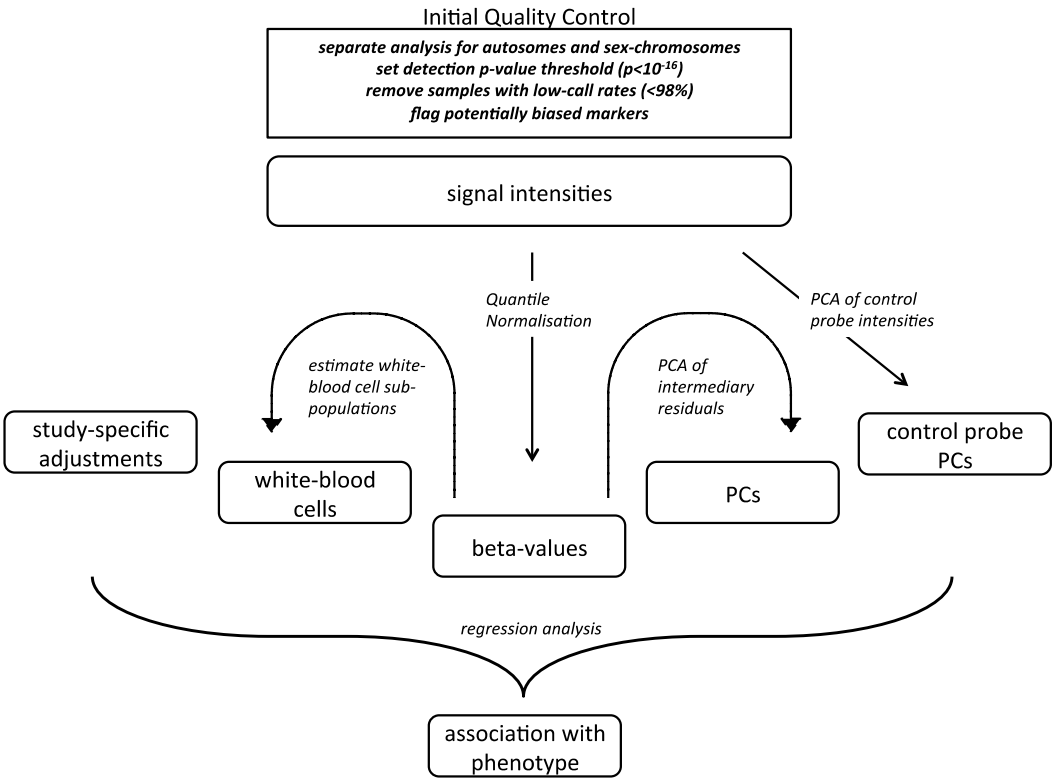


Figure 4.29: Schematic diagram of the workflow: Signal intensities were filtered for QC criteria followed by QN (I) and conversion of normalised intensities to beta values (top). A PCA is performed on the control probe intensities to derive control probe PCs (right), as well as estimated white blood cell proportions (middle left). These are then used to adjust the beta values, which in turn are the basis for the main data PCs (middle right). In the phenotypic association test (bottom), beta values are then corrected for white-blood cell proportions, control probe PCs, study-specific variables (left) and main data PCs.

4.4.1 Power to detect spiked-in Signals

To assess the performance of the above adjustments in an analysis pipeline, a simulation strategy was devised. This analysis focuses on single marker comparisons to identify differentially methylated CpG-sites, rather than a multi-marker approach to avoid the regional biases introduced. Therefore, I only considered other published single-marker approaches for comparison.

The simulated data sets were generated by Dr Benjamin Lehne as described in Chapter 2.4 (p. 39). Briefly, for each permutation, the beta values of 100 randomly selected markers were increased in samples with a case label by a predefined

increment (“spike-in”). Each spiked dataset is then processed by the complete analysis pipeline and at the different stages the ability to identify the “spiked” methylation markers based on their respective ranks is assessed. To compare the performance of these adjustments to other pipelines, the analysis of each spiked dataset was repeated with existing EWAS analysis pipelines summarised in **Table 4.4**.

Table 4.4: Comparison of available pipelines and computational requirements

Name	Preprocessing	Normalisation	Differential methylation analysis	Computational Requirements	Reference
Illumina Genome Studio	Background subtraction, detection p<0.05	Mean adjustment; normalization to internal controls	Illumina custom model	Only runs on MS windows	[191]
Minfi	Illumina control normalization; background subtraction	SWAN	Linear regression for continuous, F-test for categorical	1.9 TB RAM, 1.5 hours	[215]
IMA	Peak correction; detection p<10-5, sample call-rate>75%	Quantile normalisation	Linear regression (limma)	Computationally not tractable for >2,600 samples	[281]
Lumi	Colour bias; background correction	Quantile normalisation (M+U)	Linear regression (limma)	Computationally not tractable for >2,600 samples	[291]
Champ	SVD (visual)	BMIQ + ComBat	Linear regression (limma); lasso DMR hunter	Computationally not tractable for >2,600 samples	[278]

QN(I) yields the largest improvement and notably increases the proportion of spiked markers amongst the top 100 ranked markers (**Figure 4.30A**). For instance, at a spike-in corresponding to 15% of the SD of a marker's beta value 58% of spiked markers fall within the top 100. When QN(I) is performed, this proportion increases to 77% of markers. Adjustment for linear predictors further improves performance (**Figure 4.30A**). When adjusting for all linear predictors, 83% of spiked markers are ranked amongst the top 100 (at a spike-in of 15% SD). For spike-ins greater than 25% (SD of beta-value) close to 100% of spiked markers are amongst the top 100 irrespective of the level of adjustment.

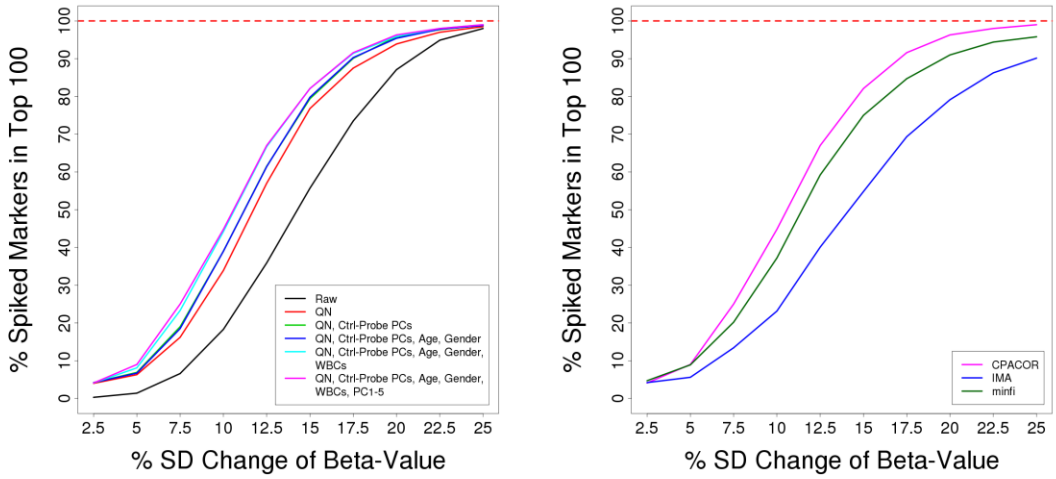


Figure 4.30: Simulation analysis of the pipeline performance. For the permutation test of 100 sets of 100 randomly selected markers were spiked with set increments in the cases and the proportion of these markers in top 100 was recorded. A: The proportion of spiked markers achieving top 100 ranks is increased by each of the stages of the workflow. B: Comparison of the pipeline with other published methods.

Most of the published methods could not be completed using datasets of >2,000 samples, even on a dedicated high-performance computing cluster with 2TB of RAM (**Table 4.4**). In contrast, the approach developed here achieves improved computational performance through parallelisation. In the two cases where

implementation was possible with a smaller data set of 500 samples, the default pipelines and single-marker association routines performed inferior to the pipeline developed here (**Figure 4.30B**).

These results further support the hypothesis that QN(I) and adjustment for the relevant linear predictors substantially improve the data quality and signal to noise ratio. Although different methylation studies may require different approaches to analysis, results from the spiked data sets indicate a significantly better performance than published methods for single-marker analysis.

4.5 Discussion

The emergence of the Illumina 450k methylation array now enables investigation of the relationships between DNA methylation and phenotype in population studies. In this work, an EWAS analysis pipeline was developed based on data from the 450k array. Main findings include that the default detection p-value is insufficiently stringent to prevent spurious results, an optimal approach to data normalisation, and a new method for dealing with technical bias. Permutation testing shows the importance of accounting for biological confounders and the null hypothesis is established after adjustment for these factors. Based the results shown in Chapter 4.3.7 (p. 114), I adopted an epigenome-wide significance threshold of $p<10^{-7}$ that is consistent with Bonferroni correction for the analyses shown in the following chapters. I further showed that, in the setting of single-marker association tests, the approach significantly outperforms existing methods for identification of true association. Furthermore, the pipeline is scalable and, unlike many existing methods, capable of handling large-scale datasets involving several thousand samples. In the following chapter, I will present the results of running the pipeline in the setting of an EWAS for BMI.

Chapter 5 An Epigenome-wide Association Study of DNA methylation with Body Mass Index

5.1 Introduction

While the previous chapter was concerned with overcoming methodological concerns with the analysis of methylation data on the 450k array, this chapter explores the use the array technology in a discovery study. This means using EWAS to identify the perturbations in DNA methylation to provide new insights into the pathways linking obesity with its consequences. The data presented represent results from the EpiMigrant BMI study, which includes meta-analysis with samples from three other studies that all adopted the same pipeline (5,387 samples overall). To confirm the primary results, I carried out replication analyses with data from 4,874 independent samples from nine different studies.

5.1.1.1 Study Design

A typical EWAS design that directly reflects GWAS involves unrelated individuals recruited in a cross-sectional study or on the basis of their phenotype (cases and controls).[192] Many case–control samples are already available, in some cases genotype and expression data already exist that can be integrated.[232] However, as has been shown in the previous chapter, methylation can change with aging and with disease onset. In addition, it is difficult to determine causality with a retrospective study.

On the other hand, a longitudinal cohort design follows cross-sectionally sampled individuals for many years, recording phenotypic and disease onset as well as taking biological samples at certain stages. Compared to retrospective designs, these studies are expensive to establish, as they follow individuals over many years.[292] The advantages of this design are the reduction of biases due to participant recruitment

and measurements of risk factors.[192] Additionally, longitudinal studies can help establish causal and consequential epigenetic variants, and changes of disease associated variants with time and environmental changes.[293]

5.1.1.2 Confounding Factors

While GWAS focus on genetic variants, which are fixed and generally cannot be changed during lifetime, EWAS face the problem of confounding through factors affecting both the phenotype and CpG methylation. This could lead to false positives, as well as inflated effect sizes. In addition to those factors identified in Chapter 4.3.7 (p.114), others are smoking behaviour, physical activity and alcohol intake, which need to be adjusted for in the regression analysis. Again, parallels can be drawn to GWAS in the sense that EWAS also employ microarrays such as the Illumina 450k array that measure a large number of sites. Therefore, principal components, as well as a correction for genome-wide inflation can be utilised.

5.1.1.3 Replication Testing

Replication testing has been shown to be an important part of EWAS to detect and prune false positives from the set of primary results.[294] The design in this study, where testing was carried out in independent individuals and by independent researchers, has the benefit of addressing technical errors and study design flaws, as well as providing additional power for the detection of true signals. An important aspect in the selection of replication studies are similar phenotype definitions to the primary study, as then results could directly be compared and meta-analysed.

5.1.2 Aims

The biological mechanisms underlying adiposity and its resulting metabolic disturbances remain poorly characterised. DNA methylation plays a key role in regulation of gene expression and phenotypic patterns.[295] In this chapter, I set out to apply the methods developed in Chapter 4 (p. 73) to the South Asian EpiMigrant

samples (Chapter 2.1.1, p. 33), and find associations with obesity (here: BMI). For the obesity association, I shared the methods with analysts working on samples from two other cohorts, KORA[296] and EPICOR, which allowed me to perform a meta-analysis for BMI between our data sets. I then performed an analysis for functional enrichment of the hits, pathway analyses, as well as association with clinical phenotypes, which were then meta-analysed with the KORA samples. Additionally, cross tissue analyses and follow-up experiments were performed by the analysis group, which I will also briefly describe here as they further characterise the identified hits.

5.2 Materials and Methods

5.2.1 Discovery Samples

Epigenome-wide association was carried out in four population cohorts comprising 5,387 individuals of European (N=2,707) or Indian Asian (N=2,680) ancestry. DNA methylation was determined by the 450k array. The European cohorts are KORA F3 and KORA F4 (**Table 5.1**), analysed by Simone Wahl, Helmholtz Institute Munich, and EPICOR, analysed by Giovanni Fiorito, University of Torino. The pipeline was implemented by the external analysts following the same procedures as outlined in Chapter 5.2.2 (p. 130), and summary statistics were provided for all analysed CpG markers.

Table 5.1: Participant characteristics for the primary cohorts

	KORA F3	KORA F4	EPICOR
N	484	1709	514
Country	Germany	Germany	Italy
Ethnicity	European	European	European
Study Design	Cohort	Cohort	Myocardial Infarction Case/Control
Age (yrs)	53.2 (9.6)	61.0 (8.9)	51.9 (7.5)
Sex (M)	52.1%	48.9%	65.8%
Fasting glucose (mmol/L)	5.8 (1.4)*	5.6 (1.1)	5.6 (1.5)
HbA1c (%)	5.3 (0.5)	5.6 (0.6)	-
Fasting insulin (IU/L)	---	8.9 (24.9)	10.7 (8.5)
% Fasting	9.5%	100%	62.8%
Weight (kg)	78.0 (15.1)	79.2 (15.3)	73.9 (12.3)
Height (cm)	169.1 (9.3)	167.8 (9.2)	166.0 (9.0)
Body mass index (kg/m^2)	27.2 (4.6)	28.1 (4.7)	26.8 (3.8)
Waist circumference (cm)	---	---	91.2 (11.4)
Waist-hip ratio	0.89 (0.09)	0.89 (0.09)	0.90 (0.08)
Alcohol (g/d)	16.1 (19.7)	15.6 (20.4)	18.2 (19.6)
LDL cholesterol (mmol/L)	3.4 (0.9)	3.6 (0.9)	3.9 (1.0)
HDL cholesterol (mmol/L)	1.5 (0.5)	1.5 (0.4)	1.5 (0.4)
Total cholesterol (mmol/L)	5.7 (1.0)	5.7 (1.0)	6.1 (1.2)
Triglycerides (mmol/L)	1.9 (1.4)	1.5 (1.1)	1.7 (1.1)
C-reactive protein (mg/L)	3.5 (4.3)	2.5 (5.1)	2.1 (2.7)
Systolic BP (mmHg)	128.7 (18.2)	124.8 (18.7)	138.0 (18.9)
Diastolic BP (mmHg)	82.8 (10.6)	76.1 (10.0)	85.4 (9.5)
HT**	45.00%	48.40%	72%
Treated HT	22.70%	37.30%	18%
CHD***	---	---	31%
T2D****	6.00%	9.00%	0.0%
Physically Active	49.8%	57.4%	78.99%
Smoking			
Never smoked	50.2%	43.8%	34.5%
Ex-smoker	0.0%	42.8%	30.8%
Current smoker	49.8%	14.5%	34.9%

Values indicate mean (SD) for each trait. *both fasting/non-fasting subjects, **HT (Hypertension): SBP ≥ 140mmHg, or DBP ≥ 90mmHg, or who were taking anti-hypertensive or blood pressure-lowering, ****T2D (Type 2 Diabetes): physician diagnosis or HbA1c>6.5%

5.2.2 Statistical analyses

Methylation of genomic DNA was initially quantified as described in Chapter 2.1.2 (p. 35). Bead intensities were retrieved using the minfi software package using background correction, and a detection p-value of < 0.01 and bead count >= 3. Signals below the threshold were excluded from the analysis.

I analysed the epigenome-wide data using minfi and custom R scripts as described in Chapter 4.4 (p. 120). Marker intensities were separately quantile normalised by probe type and colour channel and for analysis (Chapter 4.3.2 p. 95). I then used the epigenome-wide methylation values to impute 4 lymphocyte subsets (Chapter 4.3.5 p. 103).[270] To adjust for batch effects, I included the first 20 PCs of the control probe intensities. Single marker tests were performed using linear regression to examine the association of each autosomal CpG site with BMI as the outcome, adjusted for age gender, age, smoking status, physical activity index and alcohol consumption (Chapter 4.3.7 p. 114). The first 20 control probe principal components from the 450K array, total WBC count and imputed white cell subsets as covariates in the regression models. Finally, I corrected association results for the genomic control inflation factor (GC_{in}). Markers on the sex chromosomes were tested similarly for association with BMI, but separately in men and women. Analogous methods were used by Simone Wahl and Giovanni Fiorito to analyse the KORA and EPICOR samples (**Table 5.1**).

5.2.3 Meta-analysis

The METAL software (http://www.sph.umich.edu/csg/abecasis/Metal/) version 2011-03-25 was used to perform an inverse-variance balanced meta analysis of association test results from the four discovery cohorts (**Table 5.1**), which use the regression coefficients and standard errors (controlled for GC_{in})[297] from each

individual study. The resulting p-values where then corrected for in a second round of genomic control (GC_{out}). A heterogeneity test (Cochran's q test) was performed to assess whether effect sizes are homogeneous across the cohorts. The threshold for epigenome-wide significance was $p<1x10^{-7}$, as permutation testing supported (Chapter 4.3.5, p. 103). As secondary analyses, the relationship between BMI and DNA methylation amongst the 11,233 X-chromosomal and 417 Y-chromosomal CpG sites was investigated.

5.2.4 Replication Samples

I then carried out a replication analysis in a second batch of 665 samples from the LOLIPOP cohort, using the same analysis workflow as outlined in Chapter 4 (p. 73). In addition, summary statistics (mean (SD) methylation, effect (SE), call rate, p-value) for markers significant in the discovery study were provided from analysts working on samples from 8 different cohorts. The cohort participant characteristics are provided in **Table 5.2,** and the methods and analysts are given in **Table 5.3**. I performed a meta-analysis of all nine sets of samples as described in Chapter 5.2.3 (p. 130).

Table 5.2: Participant characteristics for the replication cohorts

	EGCUT Asthma	EGCUT CTG	ALSPAC
N	173	96	701
Country	Estonia	Estonia	UK
Ethnicity	European	European	European
Study Design	Asthma Case/Control	young vs old	Prospective birth cohort
Age (yrs)	26.2 (7.0)	53.0 (23.7)	47.8 (4.3)
Sex (M)	35.8%	50%	0%
Fasting glucose (mmol/L)	---	5.2 (0.9)	5.3 (1.1)
HbA1c (%)	---	---	
Fasting insulin (IU/L)	---	8.9 (6.4)	5.7 mu/L (5.4)
% Fasting	---	---	100.00%
Weight (kg)	68.7 (13.2)	77.8 (17.1)	---
Height (cm)	173.0 (10.2)	170.5 (10.8)	---
Body mass index (kg/m^2)	22.8 (3.0)	26.7 (5.1)	26.6 (5.3)
Waist circumference (cm)	78.2 (10.0)	91.0 (15.2)	84.6 (12.5)
Waist-hip ratio	0.80 (0.08)	0.87 (0.10)	0.81 (0.07)
Alcohol (g/d)	2.3 (4.7)	3.8 (9.6)	---
LDL cholesterol (mmol/L)	---	3.3 (1.0)	2.9 (0.8)
HDL cholesterol (mmol/L)	---	1.6 (0.4)	1.5 (0.4)
Total cholesterol (mmol/L)	---	5.7 (1.1)	4.9 (0.8)
Triglycerides (mmol/L)	---	1.4 (1.0)	1.0 (0.5)
C-reactive protein (mg/L)	---	1.9 (2.1)	2.1 (4.4)
Systolic BP (mmHg)	116.7 (12.0)	129.5 (19.0)	117.3 (11.5)
Diastolic BP (mmHg)	73.2 (9.4)	79.9 (10.3)	72.2 (10.3)
HT**	10.4%	50.5%	6.80%
Treated HT	1.2%	41.7%	---
CHD***	0%	13.5%	---
T2D****	0%	4.2%	---
Physically Active	74.57%	63.5%	N/A
Smoking			
Never smoked	100.0%	62.5%	57.70%
Ex-smoker	0.0%	24.0%	34.30%
Current smoker	13.5%	8.00%	9.5%

	Twins UK	The Rotterdam Study	EpiMigrant Replication
N	338	731	656
Country	UK	The Netherlands	UK
Ethnicity	European	European	South Asian
Study Design	Population/Twins	Population	Prospective T2D
Age (yrs)	54.6 (8.9)	59.9 (8.2)	53.8 (10.2)
Sex (M)	0.0%	45.90%	42.80%
Fasting glucose (mmol/L)	5.1 (0.9)*	5.6 (1.18)	5.0 (0.5)
HbA1c (%)	---	---	5.6 (0.4)
Fasting insulin (IU/L)	8.4 (7.4)	14.2 (10.8)	---
% Fasting	97.60%	100%	100%
Weight (kg)		80.7 (16.2)	72.5 (13.5)
Height (cm)	26.9 (5.0)	170.7 (10.0)	163.6 (9.1)
Body mass index (kg/m^2)	79.7 (10.6)	27.6 (4.6)	27.0 (4.4)
Waist circumference (cm)	0.80 (0.10)	93.7 (12.9)	94.5 (11.2)
Waist-hip ratio	0.79 (0.06)	0.87 (0.08)	0.93 (0.09)

Alcohol (g/d)	---	18.2 (11.3)	3.0 (8.6)
LDL cholesterol (mmol/L)	3.4 (1.0)	3.5 (0.98)	3.1 (0.9)
HDL cholesterol (mmol/L)	1.5 (0.4)	1.4 (0.41)	1.5 (0.4)
Total cholesterol (mmol/L)	5.6 (1.2)	5.6 (1.07)	5.2 (0.9)
Triglycerides (mmol/L)	1.3 (0.8)	1.5 (0.87)	1.4 (0.7)
C-reactive protein (mg/L)	---	2.7 (4.7)	---
Systolic BP (mmHg)	124.8 (16.4)	134.4 (19.8)	127.1 (18.4)
Diastolic BP (mmHg)	77 (10.1)	82.8 (11.3)	76.9 (10.6)
HT**	-	0.53%	26.20%
Treated HT	-	0.26%	---
CHD***	0.00%	5.90%	0%
T2D****	0.00%	8.50%	0%
Physically Active	N/A	---	41.50%
Smoking			
Never smoked	57.4%	29%	90.6%
Ex-smoker	33.1%	44%	4.5%
Current smoker	27%		18.40%

	LifeLines Deep	Leiden Longevity	Rotterdam Study 2
N	752	665	762
Country	The Netherlands	The Netherlands	The Netherlands
Ethnicity	European	European	European
Study Design	Population based cohort	Population based cohort	Population based cohort
Age (yrs)	45.5 (13.3)	58.9 (6.6)	67.7 (5.9)
Sex (M)	42.28%	47.26%	42.60%
Fasting glucose (mmol/L)	5.5(0.4)	---	---
HbA1c (%)	4.9(0.6)	---	---
Fasting insulin (IU/L)	---	---	---
% Fasting	99.60%	---	---
Weight (kg)	77.9 (14.1)	76.1 (13.1)	79.8 (14.1)
Height (cm)	175.2 (8.9)	172.6 (8.8)	169.3 (9.2)
Body mass index (kg/m^2)	25.4 (4.2)	25.46 (3.5)	27.8 (4.2)
Waist circumference (cm)	96.3 (10.1)	---	---
Waist-hip ratio	0.92 (0.07)	---	---
Alcohol (g/d)	---	---	---
LDL cholesterol (mmol/L)	3.1 (0.9)	---	---
HDL cholesterol (mmol/L)	1.5 (0.4)	---	---
Total cholesterol (mmol/L)	5.1 (1.0)	---	---
Triglycerides (mmol/L)	1.1 (0.9)	---	---
C-reactive protein (mg/L)	---	---	---
Systolic BP (mmHg)	119.5 (13.5)	---	---
Diastolic BP (mmHg)	71.1 (9.2)	---	---
HT**	8.37%	---	---
Treated HT	---	---	---
CHD***	---	---	---
T2D****	---	---	---
Physically Active	---	---	---
Smoking			
Never smoked	47.30%	31.6	---
Ex-smoker	34.30%	55.5	---
Current smoker	12.6	100%	

Values indicate mean (SD) for each trait. *both fasting/non-fasting subjects, **HT (Hypertension): SBP ≥ 140mmHg, or DBP ≥ 90mmHg, or who were taking anti-hypertensive or blood pressure-lowering, ****T2D (Type 2 Diabetes): physician diagnosis or HbA1c>6.5

Table 5.3: Methods used for discovery and replication. EpiM: EpiMigrant, DetP: detection p-value, CR: call rate, NBeads: number of beads detecting signal, QN: quantile normalisation, PA: Physical Activity, conc: DNA concentration (after bisulphite conversion)

	EpiM 2	EGCUT Asthma	EGCUT CTG	ALSPAC	Twins UK	Rotterdam Study	LifeLines Deep	Leiden Longevity	Rotterdam Study 2
IDAT extraction	minfi	minfi	minfi	minfi	Illumina	Methylumi	Methylumi	Methylumi	Methylumi
Background correction	minfi	none	none	none	none	separate colours	separate colors	separate colors	separate colors
DetP cutoff	0.01	0.01	0.01	0.01	0.05	0.01	N/A	N/A	N/A
Sample CR threshold	none	95%	95%	95%	N/A	95	N/A	N/A	N/A
Marker CR threshold	95%	90%	90%	90%	N/A	100%	N/A	N/A	N/A
Nbeads filter	3	N/A	N/A	N/A	3	3	N/A	N/A	N/A
Normalisation	QN, 6 categories	SWAN	SWAN	QN, 6 categories	QN	SWAN	DASEN	DASEN	DASEN
All covariates used in model (technical and biological)	Age, Sex, PA, Alcohol, Smoking, WBC, 20 control probe PCs	Age, Sex, PA, Alcohol, Smoking	Age, Sex, PA, Alcohol	Age, plate, measured WBC	Age, Smoking, conc, Array, Position, Family Zygosity	Age, Sex, Array Position	Age, Sex, Smoking	Age, Sex, Smoking	Age, Sex
Analyst	Alexander Drong	Silva Kasela	Silva Kasela	Rebecca Richmond	Dr Jordana Bell	Dr Abbas Deghan	Dr Lude Franke	Dr Lude Franke	Dr Lude Franke

5.2.5 Functional Annotation Analyses

To test for functional enrichment (chromatin structure, CpG islands and relationship to genes) of the replicated CpG sites, I used annotations for each probe provided by Illumina.[191] These included whether a CpG was in a CpG island, shore (±2kb from islands) or shelf (±2kb from islands), and annotations relative to genes and their TSS's.[191] To expand the annotation, I downloaded ENCODE[129] data from UCSC for histone modification chromatin immunoprecipitation (ChIP) peaks in the GM12878 lymphoblastoid cell line (H3K4me1, H3K4me3 and H3K27Ac), which represent marks of open chromatin, as well as DNaseI Hypersensitivity Sites (EncodeRegDnaseClusteredV2).

As a background for the enrichment test, the complete set of probes on the 450k array was used. Each probe on the 450k array background was mapped to the annotation categories and overlap at each probe was recorded as a binary variable, except for genic locations, where multiple overlaps were possible for each CpG. I tested for enrichment of the replicating markers against the background using a Fisher exact test.

5.2.6 Gene Expression

Associations between DNA methylation at validated CpGs ($p < 0.05$, $n = 187$) and expression of *cis*-genes (±500kb) were investigated in peripheral blood amongst Europeans (N = 878) by Simone Wahl, Helmholtz Institute and Indian Asians (N = 907) by Dr Marie Loh, Imperial College London. A total of 4,470 transcripts were located in *cis* to these 187 CpG sites. Pairwise association testing of a transcript and a CpG site was performed using linear models with gene expression as response variable, and DNA methylation including the covariates described in Chapter 5.2.2 (p. 130), with the addition of RNA amplification batch, RNA sample storage time and RNA integrity number (RIN)[298] as covariates. Models were fit with and without including BMI as an additional covariate. The association of DNA

methylation with each respective gene was done in each ethnic group separately, followed by meta-analyses performed by Simone Wahl. A transcript was considered associated with methylation if the p-value reached 0.05/4470 = 1.1×10^{-5} (Bonferroni correction).

5.2.7 Cross-Tissue Comparison

Cross tissue patterns of methylation were investigated by Dr Marie Loh using publicly available 450k data (GSE48472) from the Gene Expression Omnibus database.[299]. This dataset comprises 41 samples from blood and 5 metabolically relevant tissues (liver, muscle, pancreas, subcutaneous fat, omentum).

5.2.8 Candidate genes

Associations of methylation with expression have been reported to typically range ~40 kb.[300] As the standard Illumina annotation often does not contain a nearby gene if the genes are far away, I compiled a list of candidate genes for the 187 replicating CpGs as the union of two gene sets:

1. All genes within a cis region (±40 kb, multiple genes per marker)

2. All local genes (up to ±500 kb) with expression associated with the marker at an adjusted p-value of $p<0.05$

Gene annotations were downloaded from Ensmbl[301] version 27 using the R package *biomaRt* and overlapped with the CpG positions as annotated in the Illumina annotation using *GenomicRanges*.

5.2.9 Pathway Analyses

The GSEA MSigDB platform (http://www.broadinstitute.org/ gsea/msigdb) was used to investigate enrichment of the set of genes against three curated pathway sets (BIOCARTA, KEGG, REACTOME). An enrichment p-value was calculated based on a hypergeometric distribution, which was subsequently adjusted at the 5% FDR level.

I carried out a GSEA for significant markers using MAGENTA[302] with GWA data for the following phenotypes: T2D (DIAGRAM)[89], Fasting Glucose, 2hr Glucose, Fasting Insulin (MAGIC)[85], LDL, HDL (GLGC)[106], BMI, Height and WHR adjusted for BMI (GIANT)[86]. Gene boundaries were defined as 100 kb upstream and 50 kb downstream of each gene and 10,000 permutations were used. A GSEA p-value was then calculated by comparing the number of genes in the gene set in the top 5% ranked genes ("observed") versus permuted gene sets of the same size.[302]

5.2.10 Clinical implications

For 1,701 KORA F4 and 2,475 EpiMigrant subjects, fasting measurements of BMI-related clinical metabolic traits were available.[296, 303] In these, it was investigated whether the identified BMI-associated CpG sites showed a correlation with clinical traits. Association testing was performed using linear models with log-transformed and standardized clinical trait (LDL cholesterol, HDL cholesterol, total cholesterol, triglycerides, fasting glucose, fasting insulin, HbA1c, systolic and diastolic blood pressure, CRP, weight, height and WHR) as the outcome and DNA methylation at the respective trait, age, sex, smoking state, physical activity, alcohol intake and the top 20 control probe PCs as covariates. Models were fit with and without including BMI as an additional covariate. Results from KORA F4 (provided by Simone Wahl) and EpiMigrant were meta-analysed with an inverse variance weighted fixed-effects meta-analysis using METAL version 2011-03-25.

5.2.11 Causality Analysis

Differential methylation at CpGs can be the cause or the consequence of a change in BMI. To assess the direction of causality for the identified CpGs, Dr Benjamin Lehne compared predicted and observed effect sizes based on multiple regression analyses. The analyses and models for both the consequential and causal cases (**Figure 5.1**) are described in detail below. Regression summary statistics are based

on meta-analysis of LOLIPOP and KORA samples (**Table 5.4**) carried out by Simone Wahl and Dr Benjamin Lehne as well as data provided by the GIANT consortium.[86]

Figure 5.1: Mendelian Randomization triangles. Based on genetic variation (SNP or Genetic Risk Score) as instrumental variables it can be determined whether CpGs are the cause or consequence of BMI.

Table 5.4: GWAS data sets used in the causal analysis

	KORA KF4	KORA KF3	LOLIPOP IA317	LOLIPOP IA610	LOLIPOP OmniX
Call rate threshold (samples)	97%	97%	95%	95%	98%
MAF threshold	1%	1%	1%	1%	1%
Hardy-Weinberg Equilibrium threshold (P-HWE)	5.00E-06	5.00E-06	1.00E-06	1.00E-06	1.00E-06
Call rate threshold (SNPs)	98%	98%	98%	98%	98%
Info-score threshold	0.5	0.5	0.5	0.5	0.5
Platform	Affymetrix Axiom	Illumina Omni Express	Illumina 317K Beadchip	Illumina 610K Beadchip	Illumina Omni Express
Genotype calling software	Affymetrix Software	Illumina Genome Studio	Illumina Genome Studio	Illumina Genome Studio	Illumina Genome Studio (Gencall module) and zCall
Imputation panel	1000g phase1	1000g phase1	1000g phase1	1000g phase1	1000g phase1
Imputation software	IMPUTE v2.3.0	IMPUTE v2.3.0	IMPUTE v2.3.0	IMPUTE v2.3.0	IMPUTE v2.3.0
Cohort-specific adjustments	none	none	PC1-5 (genetic)	PC1-5 (genetic), Coronary Heart Disease	PC1-5 (genetic)
Removal criteria	gender discordance, high or low heterozygosity, cryptic relatedness, population outliers	gender discordance, high or low heterozygosity, cryptic relatedness, population outliers	gender discordance, high or low heterozygosity, cryptic relatedness, population outliers	gender discordance, high or low heterozygosity, cryptic relatedness, population outliers	gender discordance, high or low heterozygosity, cryptic relatedness, population outliers

5.2.11.1 Causal model

1. *CpG ~ SNP + discovery covariates*

In the first step, SNPs controlling methylation at the CpG site in *cis* (±1 Mb) were detected by association testing and subsequently meta-analysed (inverse-variance weighted) across the five cohorts described above. Of the 187 CpG site associated with BMI, 175 CpGs had *cis* SNPs shared between all cohorts and no SNPs underlying the probe sequence. For each CpG site, the SNP with the lowest p-value was recorded and its respective P-value was adjusted for the number of SNPs allocated to the CpG (p_{adj}). A SNP is considered significantly associated if it passed Bonferroni threshold of $p_{adj} < 0.05$ / Number of CpGs with associated SNP.

2. *BMI ~ CpG + discovery covariates*

An association test for inverse-normalised BMI with each CpG site was carried out in order to compare effect sizes with the latest GWAS for BMI. The results were meta-analysed across the EpiMigrant and KORA cohorts.

3. *BMI ~ SNP + Age + Gender (observed)*

Regression summary statistics were retrieved from a large GWAS meta-analysis (~100,000 samples)[86] for the 175 SNPs most strongly associated with the 175 CpG sites. Effect sizes were corrected to match the effect allele of the SNPs in step 1. The significance threshold for an association with BMI was set to $p < 0.05$ / Number of CpGs with associated SNP.

4. *BMI ~ SNP (predicted)*

To compare the observed effect size with the predicted effect size under a causal model, effect sizes and standard errors (β_{pred} and SE_{pred}) are derived as follows:

$$\beta_{pred} = \beta_{CpG \sim SNP} \times \beta_{BMI \sim CpG}$$

$$SE_{pred} = \sqrt{SE_{CpG\sim SNP}^2 \times SE_{BMI\sim CpG}^2 + SE_{CpG\sim SNP}^2 \times \beta_{BMI\sim CpG}^2 + SE_{BMI\sim CpG}^2 \times \beta_{CpG\sim SNP}^2}$$

5.2.11.2 Consequential model

1. *BMI ~ GRS*

A genomic risk score (GRS) for each individual in the KORA and EpiMigrant cohort was calculated based on 32 SNPs with replicated BMI associations.[86] One SNP (rs7359397) was removed as it associated ($p < 0.05 / (187 \times 32)$) with a CpG site after adjustment for BMI. The number of copies of the risk allele (dosage) weighted by effect size of the respective SNP[86] is summed for each sample. BMI was inverse normal transformed and regression results for the five cohorts were meta-analysed (inverse-variance weighted).

2. *CpG ~ BMI + discovery covariates*

An association test for invers-normalised BMI with each CpG site was carried out as described in Chapter 5.2.11.1 (p. 141).

3. *CpG ~ GRS + discovery covariates*

Association between 187 CpGs and GRS was tested in EpiMigrant and KORA cohorts and meta-analysed (inverse-variance weighted). CpGs are considered as significantly associated with GRS if P < 0.05 / Number of CpGs.

4. Effect sizes and standard errors.

Effect sizes and standard errors (β_{pred} and SE_{pred}) were derived as described in Chapter 5.2.11.1 (p. 141).

5.3 Results

5.3.1 Epigenome-wide association testing

I first carried out a single-marker EWAS using the 2,680 EpiMigrant samples described in Chapter 2.1.1 (p. 33), which was carried out in parallel to analyses performed for KORA F3 and KORA F4 by Simone Wahl and EPICOR by Simone Wahl. QQ plots of p-value distributions of the analyses are shown in **Figure 5.2,** with genomic inflation factors (λ) ranging from 0.982 to 1.286. The results indicate a large number associations of methylation at CpG sites with BMI.

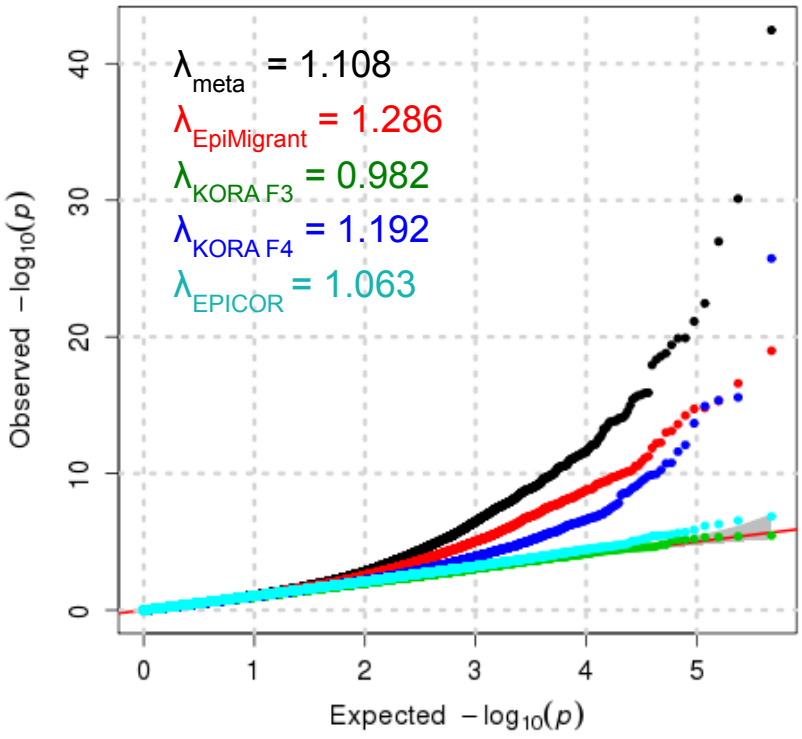


Figure 5.2: Discovery Stage association testing of BMI against methylation and meta-analysis. QQ plots show p-values for association tests of BMI against methylation for the combined meta-analysis and each study included in the discovery phase. The theoretical 95% CI is shown in grey. Black: Combined meta-analysis, Red: EpiMigrant, Green: KORA F3, Blue: KORA F3, Turquoise: EPICOR.

I then carried out a meta-analysis of the EWAS results from the individual cohorts and found 278 markers associated with BMI at $p < 1x10^{-7}$, distributed between 207

distinct genetic loci (**Appendix Table 2**). At each locus the sentinel marker was identified (CpG site with lowest p-value against BMI), and replication testing was carried out in separate samples of European and Indian Asian men and women from 6 population-based studies (N = 4,874**).** Of the 207 markers, 187 reached $p < 0.05$ in replication testing (**Figure 5.3**), and in combined analysis of discovery and replication data all these 187 markers exceeded the threshold for epigenome-wide significance by an order of magnitude (**Table 5.3, Appendix Table 2**).

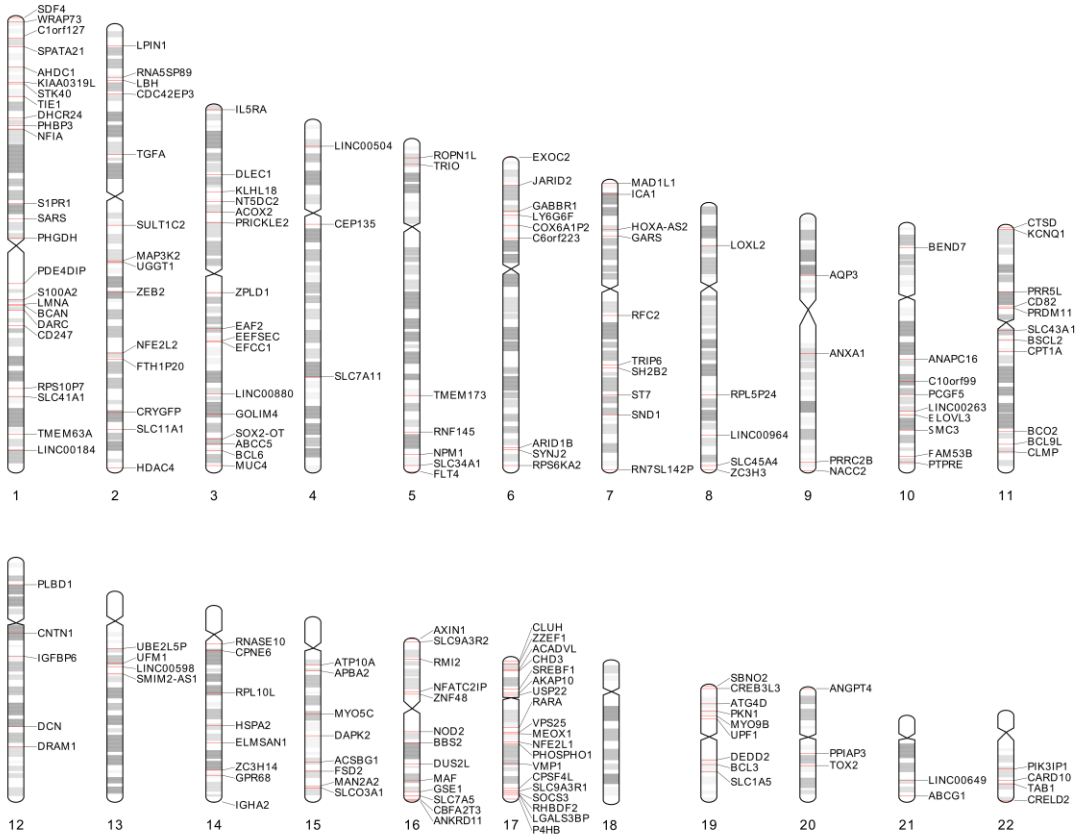


Figure 5.3: Chromosome of ideogram replicating lead markers. Karyogram of human chromosomes 1-22 showing approximate position of the 187 replicating ($p<0.05$ in the replication meta-analysis) EWAS hits for BMI.

Table 5.4: Top 20 CpGs associated with BMI.

			Discovery			Replication		
CpG	**Chr-Position**	**Nearest Gene**	**Effect (SE)**	**P-value**	**HetP**	**Effect (SE)**	**P-value**	**HetP**
cg06500161	21-43656587	*ABCG1*	34.8 (2.3)	6.0E-46	4.7E-2	19.9 (1.9)	1.1E-25	5.2E-3
cg00574958	11-68607622	*CPT1A*	-40.2 (3.1)	3.3E-33	8.2E-1	-21.3 (2.4)	2.8E-18	1.4E-1
cg12593793	1-156074135	*LMNA*	-32.0 (2.7)	3.0E-27	3.6E-1	-10.2 (1.7)	3.9E-9	1.4E-2
cg11024682	17-17730094	*SREBF1*	32.5 (3.0)	1.9E-23	7.9E-5	22.7 (2.2)	3.0E-25	6.0E-1
cg27243685	21-43642366	*ABCG1*	31.0 (3.1)	8.2E-21	3.4E-4	17.4 (2.5)	6.7E-12	8.1E-5
cg09349128	22-50327986	*CRELD2*	-24.4 (2.4)	1.7E-20	3.7E-3	-13.7 (1.9)	7.3E-13	1.2E-2
cg18181703	17-76354621	*SOCS3*	-18.0 (1.8)	2.8E-20	5.5E-1	-12.1 (1.5)	1.6E-16	4.7E-2
cg17501210	6-166970252	*RPS6KA2*	-18.4 (1.9)	8.2E-19	1.8E-3	-15.2 (1.6)	6.4E-21	3.1E-1
cg16594806	1-59473943	*PHBP3*	-24.9 (2.6)	2.3E-18	5.3E-5	-3.5 (1.4)	1.3E-2	5.2E-1
cg06164260	3-187454439	*BCL6*	-21.3 (2.3)	8.1E-18	9.5E-2	-13.0 (2.2)	4.9E-9	5.5E-2
cg14476101	1-120255992	*PHGDH*	-14.5 (1.6)	2.1E-17	5.2E-2	-8.9 (1.2)	9.4E-13	6.5E-1
cg02650017	17-47301614	*PHOSPHO1*	-28.6 (3.2)	1.9E-16	4.7E-1	-15.2 (2.7)	2.0E-8	2.2E-4
cg19750657	13-38935967	*UFM1*	23.9 (2.7)	2.2E-16	1.9E-6	10.8 (1.6)	4.7E-11	6.1E-1
cg09664445	17-2612406	*CLUH*	31.2 (3.5)	2.4E-16	5.4E-2	8.5 (2.2)	1.2E-4	5.7E-2
cg17260706	11-118782879	*BCL9L*	-28.4 (3.3)	1.2E-15	7.7E-2	-9.7 (2.1)	4.7E-6	6.0E-1
cg26403843	5-158634085	*RNF145*	10.7 (1.3)	2.9E-15	4.2E-1	7.7 (1.0)	5.1E-14	7.5E-1
cg11376147	11-57261198	*SLC43A1*	-25.3 (3.0)	5.8E-15	1.1E-4	-9.1 (2.4)	1.6E-4	7.2E-1
cg12484113	1-27898757	*AHDC1*	18.6 (2.2)	1.2E-14	1.8E-1	14.3 (2.4)	2.2E-9	4.6E-1
cg08857797	17-40927699	*VPS25*	19.8 (2.4)	3.3E-14	1.3E-5	3.9 (1.3)	2.3E-3	4.3E-2

Chr: Chromosome. Discovery Effect (SE): Effect (SE) of the CpG effect in the meta-analyses, P-value: p-value of the CpG effect in the meta-analyses, HetP: Heterogeneity p-value (Chochran's q-test) of the CpG effect in the meta-analyses

Amongst the 20 markers which did not reach $p < 0.05$ in replication testing (**Appendix Table 3)**, all 20 showed consistent direction of effect ($p < 10^{-100}$, binomial test), indicating that the majority are also likely to be real and that the absence of replication in part reflects the lower sample size in replication than discovery. There were no additional genomic regions containing DNA methylation associated with BMI at $p < 10^{-7}$ on either of the sex chromosomes. Conditional analyses suggest the presence of multiple CpG sites independently associated with BMI at 23 of the identified loci ($p < 10^{-7}$ after conditioning on sentinel CpG site at the locus **(Table 5.4).**

Table 5.5: Secondary signals and conditional analysis at loci with multiple hits.

Lead Marker	Secondary Marker	Distance	r^2	P-value lead	P-value second	P-value of second cond. lead
cg00431050	cg00547480	109769	0.61	2.88E-12	5.98E-10	2.35E-12
cg00574958	cg17058475	115	0.64	1.38E-38	3.94E-14	1.01E-01
cg00711896	cg01304182	143	0.78	1.43E-10	3.63E-09	5.74E-03
cg00973118	cg16778405	278873	-0.06	1.48E-09	8.01E-09	1.58E-09
cg01101459	cg00851028	34295	0.21	9.67E-12	1.06E-11	1.06E-09
cg01243823	cg06207201	30677	0.16	2.17E-14	1.54E-11	4.33E-08
cg01243823	cg09559780	4991	0.15	2.17E-14	8.70E-10	6.19E-05
cg02711608	cg21766592	102	0.61	1.33E-12	1.14E-09	4.23E-03
cg03318904	cg06397161	41463	-0.12	2.96E-14	5.66E-10	1.86E-09
cg03318904	cg22650271	41357	-0.21	2.96E-14	6.53E-09	1.04E-08
cg04407417	cg03747456	357476	-0.03	3.28E-09	4.83E-09	6.74E-11
cg05095590	cg08972190	264	0.47	3.64E-11	3.57E-10	2.06E-07
cg05648472	cg00417304	107908	-0.26	1.46E-12	2.44E-10	3.69E-09
cg06500161	cg01176028	3353	0.22	1.73E-53	1.07E-09	3.51E-07
cg06500161	cg27243685	14221	0.46	1.73E-53	4.32E-24	1.51E-06
cg06500161	cg10192877	14897	0.14	1.73E-53	3.77E-09	6.70E-05
cg06500161	cg01881899	3883	0.30	1.73E-53	1.73E-10	1.22E-02
cg06603309	cg02079413	262361	-0.16	2.51E-13	5.25E-10	3.51E-12
cg07021906	cg02203067	33	0.79	1.03E-11	1.40E-10	4.98E-04
cg07136133	cg00220721	66	0.74	1.58E-12	1.20E-10	5.62E-03
cg07202479	cg10636246	127189	0.30	9.39E-11	1.32E-09	2.28E-12
cg07769588	cg22898082	418806	0.21	1.66E-14	1.75E-09	1.91E-07
cg07769588	cg10819350	64	0.75	1.66E-14	5.38E-11	2.49E-03
cg08309687	cg27037013	71	0.82	9.40E-16	2.69E-10	3.96E-02
cg08726900	cg06988336	371923	-0.11	8.57E-11	6.16E-09	2.55E-09
cg08857797	cg13467672	463074	0.20	2.22E-16	2.31E-09	1.01E-11
cg08857797	cg15138339	212477	-0.06	2.22E-16	5.15E-11	1.28E-08
cg08857797	cg19618279	212471	0.19	2.22E-16	1.22E-09	1.58E-06
cg09315878	cg09276842	2446	0.70	2.18E-09	4.20E-09	2.32E-04
cg09349128	cg09182678	725	0.50	9.90E-24	3.51E-12	1.06E-07
cg09664445	cg00441709	84126	0.21	7.11E-19	2.99E-11	2.22E-11
cg11080651	cg17320136	122382	0.54	1.33E-10	5.93E-10	6.62E-12
cg11080651	cg13276570	122120	0.00	1.33E-10	3.25E-09	2.30E-10
cg11080651	cg22143698	162535	0.47	1.33E-10	4.27E-09	1.39E-09
cg11202345	cg14870271	47	0.82	4.62E-13	3.50E-12	2.97E-03
cg11202345	cg25178683	210	0.65	4.62E-13	7.62E-10	7.64E-03
cg11202345	cg17836612	300	0.61	4.62E-13	2.07E-10	1.44E-02
cg11202345	cg04927537	34	0.80	4.62E-13	8.13E-09	2.26E-02
cg11832534	cg16395997	1200	0.60	7.86E-10	5.09E-09	1.11E-04
cg11969813	cg01419914	441868	0.19	3.15E-13	5.12E-09	3.30E-09

cg11969813	cg07012687	378621	0.41	3.15E-13	8.21E-10	1.43E-07
cg11969813	cg10937494	378542	0.58	3.15E-13	2.15E-09	9.33E-07
cg11969813	cg19420720	55	0.79	3.15E-13	1.81E-12	8.07E-03
cg11969813	cg17054691	3120	0.77	3.15E-13	7.67E-09	4.46E-01
cg12593793	cg08881019	22043	0.56	1.25E-31	8.23E-09	6.78E-03
cg13123009	cg10970124	47280	0.34	1.61E-10	3.61E-09	1.16E-06
cg13781414	cg14559176	78749	0.27	3.03E-09	4.13E-09	7.83E-10
cg14476101	cg16246545	51	0.88	4.08E-20	5.11E-15	3.87E-01
cg15721584	cg09831562	370	0.48	3.53E-16	1.15E-10	9.82E-06
cg15721584	cg02417857	49	0.67	3.53E-16	4.08E-10	3.92E-03
cg18120259	cg06560379	336666	-0.04	6.46E-11	1.81E-10	5.22E-12
cg18181703	cg04610187	6173	0.37	1.83E-23	6.12E-12	5.74E-06
cg18181703	cg27637521	581	0.57	1.83E-23	4.29E-10	7.47E-03
cg18181703	cg10508317	525	0.74	1.83E-23	2.97E-11	1.11E-01
cg18219562	cg03078551	117345	-0.24	2.93E-10	1.54E-09	4.60E-14
cg18608055	cg00994936	293036	-0.35	6.48E-15	8.94E-15	4.10E-18
cg18608055	cg24843003	278681	0.12	6.48E-15	5.75E-12	1.14E-17
cg18608055	cg23866916	24872	-0.53	6.48E-15	3.36E-14	6.68E-15
cg18608055	cg07573872	4524	0.31	6.48E-15	2.30E-13	3.63E-14
cg18608055	cg12272104	293101	-0.26	6.48E-15	1.25E-09	9.78E-11
cg19748495	cg06180099	5	0.90	1.13E-10	2.84E-09	9.45E-03
cg19998073	cg00871207	1807	0.10	5.41E-11	3.12E-09	1.29E-07
cg21037180	cg25230327	233994	0.15	1.08E-09	2.62E-09	9.21E-11
cg21429551	cg23068772	101299	0.16	2.20E-11	8.40E-11	1.41E-11
cg22103219	cg23261443	338552	-0.30	1.18E-13	1.19E-09	2.70E-10
cg22700686	cg24678869	380874	0.42	4.72E-14	7.97E-13	6.88E-16
cg22700686	cg22441770	390828	-0.55	4.72E-14	9.50E-11	3.02E-13
cg22700686	cg13997435	358	0.08	4.72E-14	6.46E-11	1.08E-06
cg26952928	cg25392060	66888	0.41	1.69E-10	4.89E-09	4.05E-10
cg26952928	cg26610247	66942	0.42	1.69E-10	3.78E-09	4.17E-09
cg27087650	cg26470501	2841	0.19	9.03E-11	7.82E-10	1.74E-07

Distance: distance between the two CpG sites, r^2: correlation coefficient between the two CpG sites, P-value lead: p-value for the effect of the lead CpG site, P-value second: p-value for the effect of the secondary CpG site, P-value second cond. lead: p-value for the effect of the secondary CpG site including the lead CpG site in the model

5.3.2 Functional Genomics

I investigated the distribution of the 187 CpG sites associated with BMI in relation to core functional genomic features, including the relationship to open chromatin, location in/near genes, and the relationship to CpG islands using data from the ENCODE monocyte cell line GM12878.[304] The results show that the CpG sites associated with BMI are enriched with respect to active chromatin sites (**Figure 5.4A**), including DNase hypersensitivity sites ($p = 2.4x10^{-7}$), enhancers ($p = 3.5x10^{-9}$) and the activating histone mark H3K4me1 ($p = 3.9x10^{-12}$). There was also a significant enrichment for CpGs within gene bodies (**Figure 5.4B**, $p=2.6x10^{-4}$), but depletion at transcriptional start sites ($2.5x10^{-9}$). The CpG sites were also enriched in "open sea" locations (**Figure 5.4C** $p = 4.5x10^{-9}$) and CpG shelves, but are markedly depleted in CpG Islands ($p = 1.7x10^{-25}$).

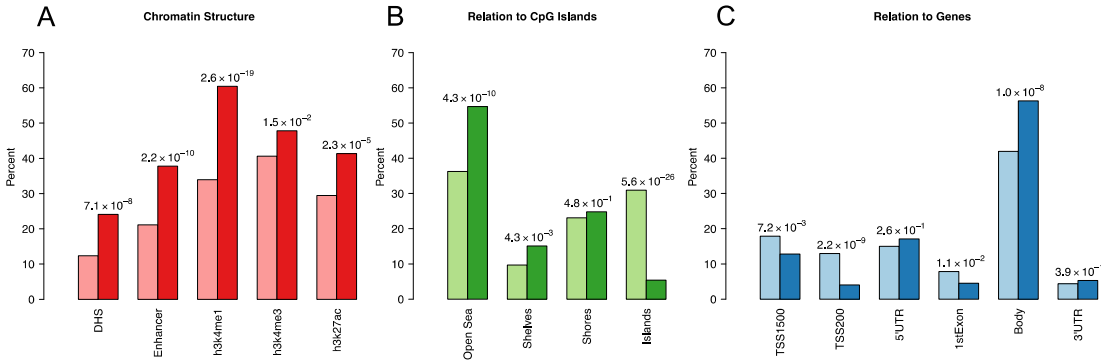


Figure 5.4: Enrichment analysis for genomic features. Light colours show percentages in the array background, whereas light colours show percentages within the 187 replicating CpG sites. A: CpG Islands (provided by Illumina[191]) B: Location with respect to genes (provided by Illumina[191]) C: DNAseI hypersensitivity sites (provided by provided by Illumina Bibiokva ref). D: Enhancers (provided by provided by Illumina[191]). E: Histone 3 Lysine 4 monomethylation (from ENCODE[305]) F: Histone 3 Lysine 27 acetylation (from ENCODE[305])

5.3.3 Cross-Tissue Comparison

Dr Marie Loh investigated the mean methylation levels of the 187 CpGs in multiple tissues.[299] Results show that DNA methylation in blood is highly correlated with methylation in subcutaneous (r=0.61) and omental (r=0.78) fat (**Figure 5.5**). The

correlations with liver (r=0.50), muscle (r=0.37) and other metabolically relevant tissues are lower with evidence for tissue-specific patterns of methylation.

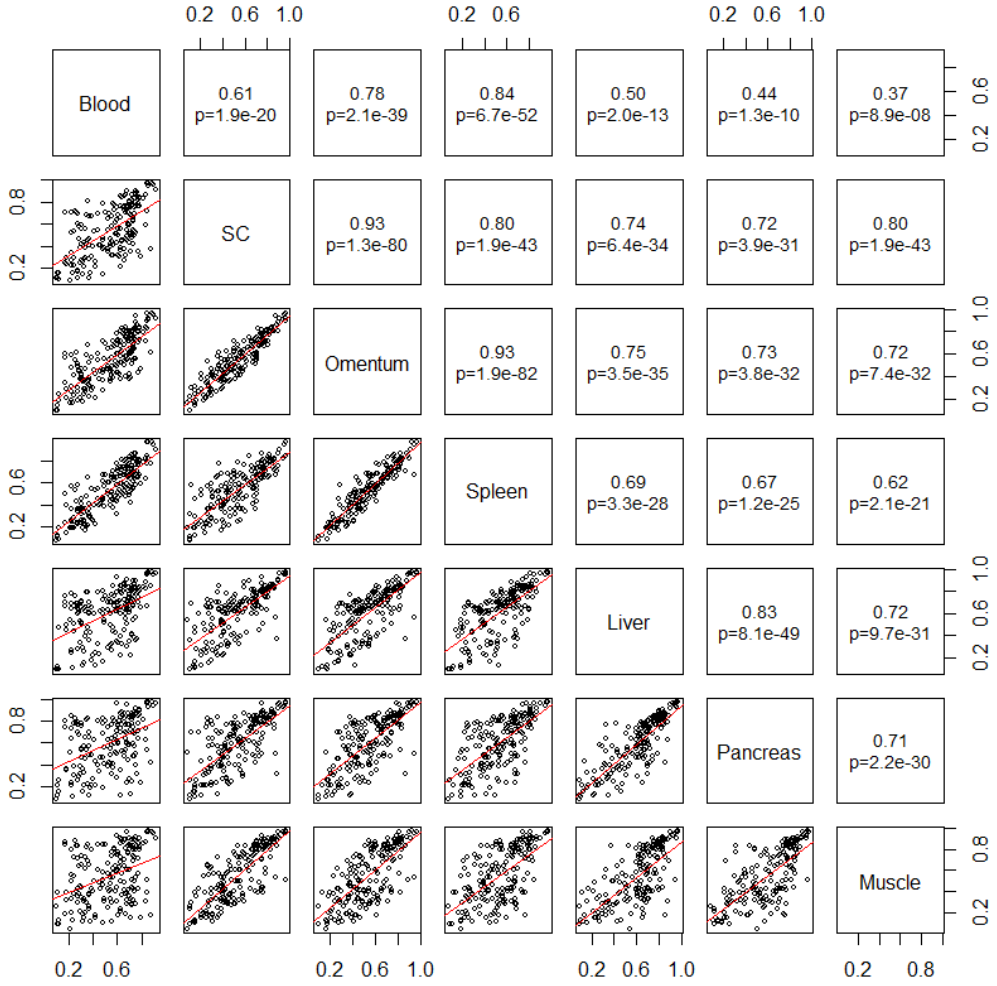


Figure 5.5: Cross-tissue comparisons of mean methylation values at the 187 CpG sites associated with BMI: At each of the 187 replicating markers, the mean methylation beta was calculated and then compared to mean methylation values in a publically available dataset of multiple tissues. For each tissue pair, the Pearson correlation coefficient for all 187 mean methylation values was calculated. The resulting scatterplots (bottom triangle) and correlation coefficients and p-values (upper triangle) are shown in the pair plot. SC: subcutaneous abdominal adipose tissue

5.3.4 Gene Expression in peripheral blood

First, the associations between DNA methylation at the 187 validated CpGs and expression of *cis*-genes (±500 kb) in peripheral blood in a subset of samples

amongst Europeans (N = 878) and Indian Asians (N = 907) were investigated. The association of DNA methylation with each respective gene was done in each ethnic group separately, followed by meta-analyses. A total of 44 transcripts of 38 annotated genes were associated with DNA methylation at $p < 0.05$ after Bonferroni correction for the 5,551 methylation marker-gene transcript statistical tests performed (**Table 5.6**). The majority of the associations were negative. The strongest signals were observed for cg09315878 with *SDF4* transcription ($p = 2.4x10^{-84}$), cg14476101 with *PHGDH* transcription ($p = 1.6x10^{-64}$) and cg09152259 with *MAP3K2* transcription ($p = 3.6x10^{-67}$).

Table 5.6: Associations of 187 BMI CpGs with expression at nearby (±500 kb)

CpG ID	Chr	Transcript	Effect	SE	P-value	HetP
cg09315878	1	SDF4	-4.58	0.23	7.23E-86	9.67E-01
cg09152259	2	MAP3K2	-3.41	0.20	1.55E-67	1.69E-01
cg14476101	1	PHGDH	-2.69	0.16	1.04E-64	4.82E-03
cg23032421	3	IL5RA	-3.63	0.26	2.69E-45	4.60E-10
cg06500161	21	ABCG1	-4.43	0.35	1.44E-36	1.74E-01
cg01798813	17	SPNS3	-4.97	0.43	1.98E-30	1.53E-05
cg00863378	16	BBS2	-2.28	0.21	2.48E-27	4.41E-01
cg00863378	16	BBS2	-2.17	0.22	2.56E-22	2.28E-01
cg23032421	3	IL5RA	-2.26	0.26	1.07E-18	7.07E-05
cg06500161	21	ABCG1	-2.35	0.27	1.19E-18	6.03E-10
cg24469729	7	HOXA5	-1.61	0.19	7.40E-18	2.55E-02
cg11202345	17	LGALS3BP	-1.01	0.12	8.57E-17	3.47E-02
cg10922280	16	DPEP2	-2.61	0.33	1.68E-15	2.21E-12
cg09315878	1	TNFRSF18	-0.65	0.08	7.17E-15	3.56E-12
cg09613192	2	UBE2E3	-1.10	0.16	5.29E-12	1.64E-01
cg00711896	16	C16ORF93	-1.90	0.28	1.79E-11	2.01E-02
cg10922280	16	DUS2L	-2.10	0.31	2.17E-11	1.36E-04
cg24469729	7	SKAP2	-1.88	0.29	1.31E-10	4.50E-01
cg00973118	16	TMEM8A	1.96	0.31	2.49E-10	8.74E-01
cg14020176	17	CD300LF	-1.84	0.30	1.15E-09	1.08E-01
cg04232128	5	NA	2.83	0.47	1.45E-09	8.48E-01
cg17901584	1	DHCR24	-0.57	0.10	2.73E-09	2.65E-04
cg13591783	9	ANXA1	-2.13	0.38	1.41E-08	5.62E-03
cg24469729	7	SKAP2	-2.03	0.36	1.70E-08	5.35E-02
cg01798813	17	CTNS	-1.88	0.34	5.19E-08	7.83E-05
cg16611584	17	AKAP10	-0.51	0.10	2.24E-07	4.64E-01
cg07037944	15	DAPK2	1.76	0.35	3.51E-07	4.79E-01
cg11024682	17	SREBF1	-1.33	0.27	5.97E-07	1.22E-01

cg08813944	17	CDC42EP4	-1.08	0.22	7.98E-07	3.92E-01
cg08548559	22	SELM	-0.79	0.16	9.12E-07	6.35E-01
cg07136133	11	PRR5L	-1.33	0.27	9.16E-07	2.77E-01
cg03957124	6	NA	4.43	0.92	1.40E-06	8.23E-01
cg19589396	8	AZIN1	0.88	0.19	2.37E-06	7.28E-02
cg11183227	15	FES	-1.18	0.26	3.78E-06	1.62E-01
cg25096107	14	NA	2.97	0.64	3.98E-06	9.39E-01
cg07769588	19	S1PR5	-0.88	0.19	4.06E-06	7.93E-01
cg00711896	16	LOC606724;CORO1A	-1.37	0.30	4.39E-06	8.44E-01
cg19589396	8	AZIN1	1.50	0.33	4.63E-06	2.89E-03
cg13123009	6	HSPA1L	-1.12	0.25	5.38E-06	3.06E-01
cg10922280	16	DDX28	-1.17	0.26	6.76E-06	1.10E-02
cg25096107	14	KIAA0125	-1.87	0.42	6.99E-06	2.63E-02
cg00094412	6	GABBR1	-1.14	0.25	7.50E-06	1.79E-01
cg04232128	5	CXXC5	2.90	0.65	7.67E-06	2.01E-01
cg21108085	11	TSPAN18	-1.77	0.40	7.80E-06	1.44E-03

Chr: Chromosome**,** Effect: beta coefficient of the CpG site, SE: SE of the CpG site **,** P-value: P-value of the CpG site, HetP: Heterogeneity p-value (Chochran's q-test) of the CpG site

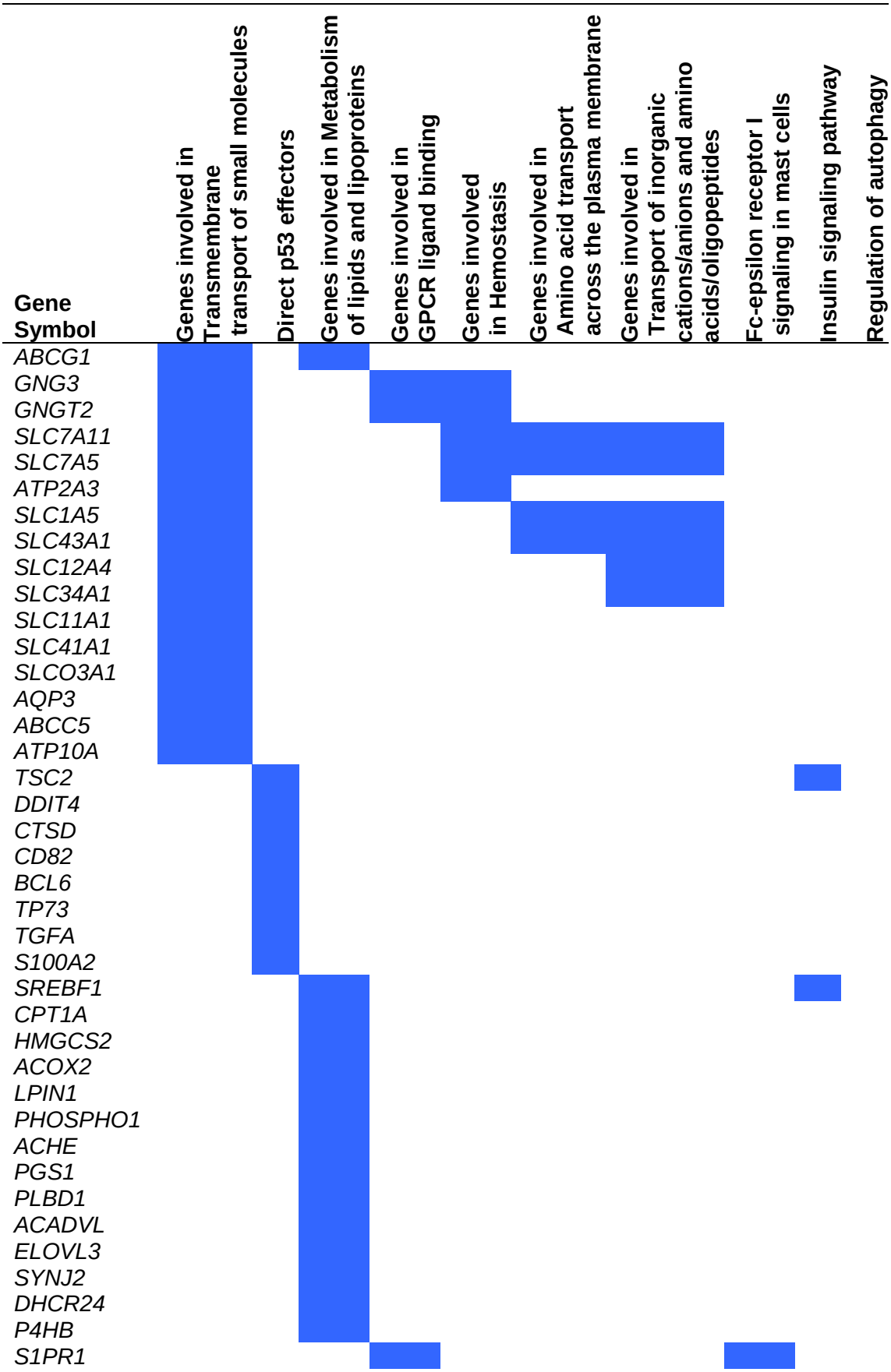
To assess whether the findings in peripheral blood might reflect processes in metabolically relevant tissues, the relationship between DNA methylation and gene expression in liver (N = 70) was examined. Amongst the 38 genes showing association of DNA methylation with expression in blood, 15 genes showed nominal association ($p < 0.05$) between methylation and expression in liver, and 5 were associated at $p < 0.05$ after correction for multiple testing ($p = 0.01$ compared to expectations under null hypothesis).

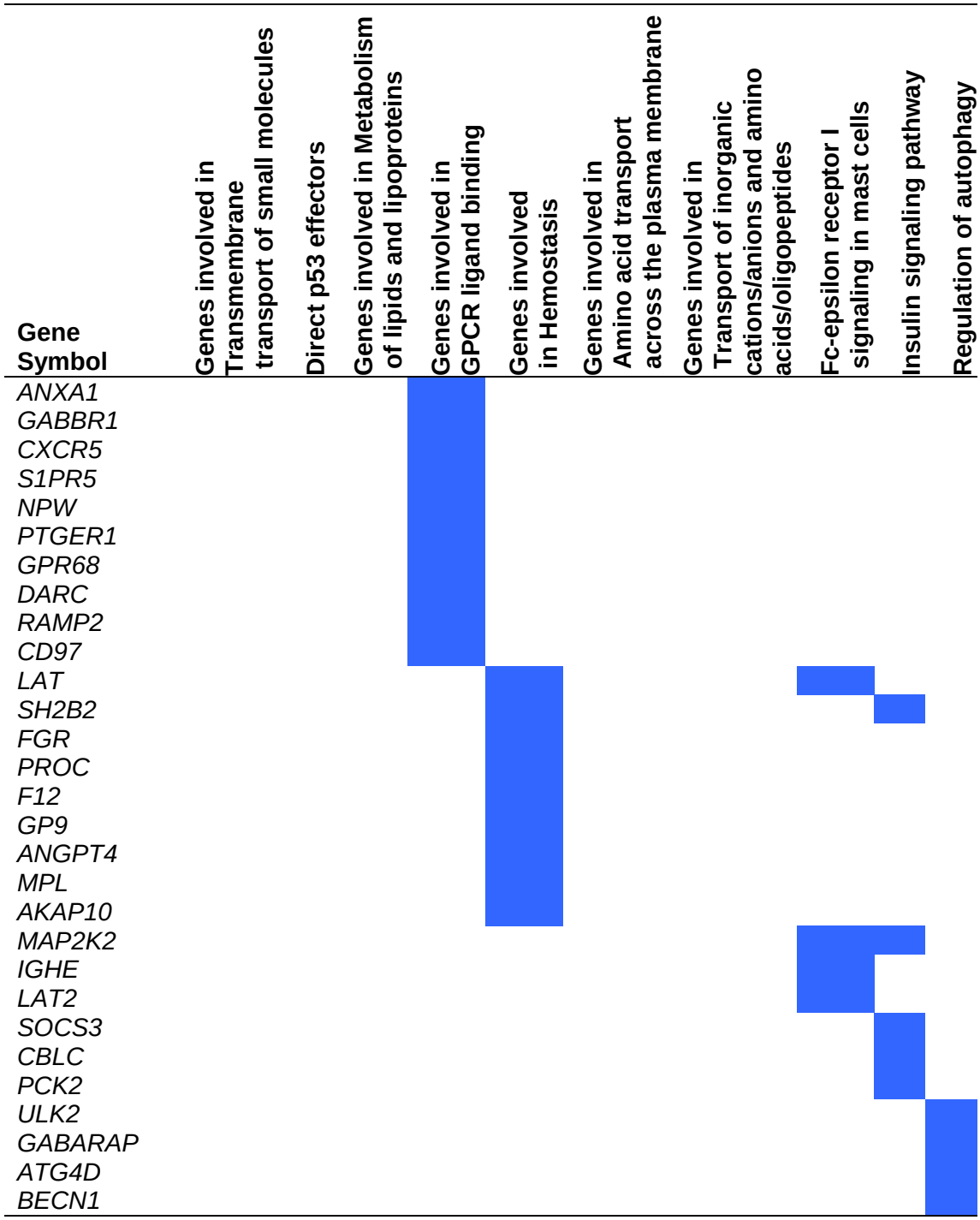
5.3.5 Candidate genes at the loci associated with BMI

Potential candidates for genes linked to the differential methylation at the 187 genetic loci were selected using the following criteria: i. nearest gene to the sentinel methylation marker, and another gene within 40 kb (529 genes. This distance was selected based on a previous report that methylation typically associates with gene expression over these distances[299]; ii. genes within 500kb of the sentinel methylation marker showing association gene expression with methylation at $p < 0.05$ (38 genes). This yielded a combined list of 546 unique genes.

The GSEA MSigDB platform was used to investigate enrichment of the set of 546 genes against three curated pathway sets. GSEA revealed enrichment for genes involved in metabolic signalling (lipid metabolism, insulin signalling and PPAR signalling); transmembrane transport (solute carrier proteins, binding cassettes) and in haemostasis and autophagy (**Table 5.7**).

Table 5.7: Pathway enrichment analysis of genes related to 187 BMI CpG sites





Blue bars: genes overlapping pathway, P-value: enrichment p-value for the pathway (hypergeometric test), FDR q-value: P-values corrected at the 5% FDR level

5.3.6 Clinical implications

Next, I investigated whether the DNA methylation might mediate the relations of adiposity with its adverse metabolic outcomes. First, the relationship of the 187 sentinel CpG sites with 13 clinical traits associated with adiposity was tested. Of 2,431 methylation-clinical trait pairs tested, 879 were significant at a $p < 0.05$ after Bonferroni correction for the number of tests performed (2,431), representing a 7-fold enrichment compared to expectations under the null hypothesis ($P<10^{-100}$; binomial, **Figure 5.6**). Full association testing results are shown in **Appendix Table 3**.

I then tested whether these CpG sites accounted for a part of the association of BMI with the clinical phenotype. The association of BMI and clinical trait was tested with and without adjustment for methylation at the 187 sentinel CpG sites. Then, the change in beta-coefficient after adjustment for methylation was used to quantify the possible contribution of methylation to the relationship between BMI and phenotype. P-values were calculated by 10,000 bootstrap samples with permutation of the CpG values. The proportion of the relationship between BMI and phenotype ranged from 14.5% (systolic blood pressure; $p = 2.3x^{-4}$) to 68.4% (triglycerides; $p = 2.9x10^{-16}$).

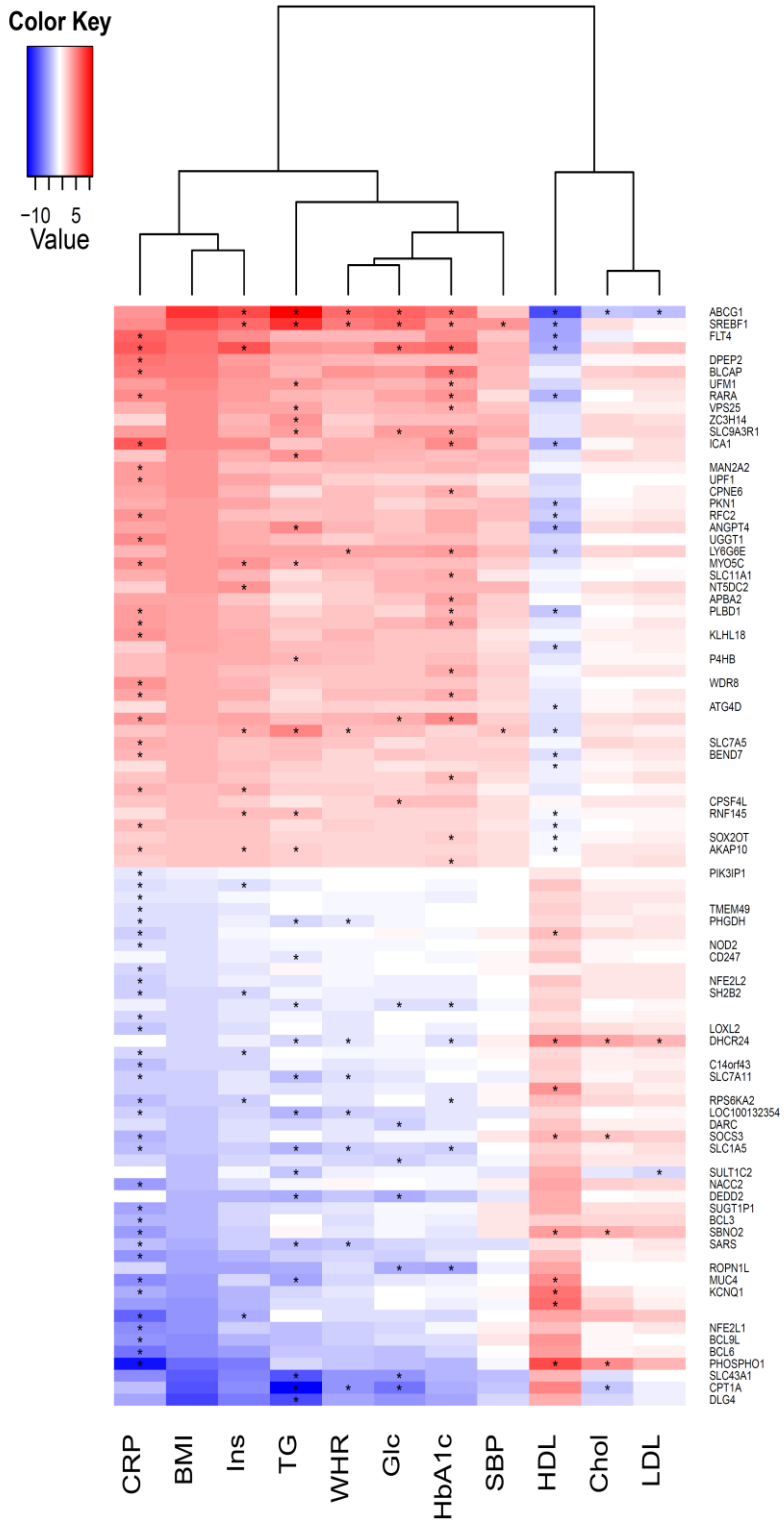


Figure 5.6: Association with clinical phenotypes: The 187 replicating markers were tested for association with 13 clinical phenotypes using a model with the same covariates as in the main discovery studies including BMI. The individual results were then meta-analysed between the four discovery cohorts and p-values are presented on a heat map. Black stars represent results significant at $p<2.1x10^{-5}$.

5.3.7 Causality Analysis

To establish whether the CpG associating with BMI are causal for the effect on BMI or a consequence, Dr Benjamin Lehne performed Mendelian Randomization experiments comparing predicted and observed effects for each CpG passing replication (Chapter 5.2.11 p. 138). At one CpG site (cg26663590: *NFATC2IP*), there is evidence for causality (**Figure 5.6 A**). On the other hand, when considering DNA methylation as a consequence of BMI, five CpGs pass multiple hypothesis testing correction and a strong correlation between predicted and observed effects (Pearson's R=0.79; P=6.4 x 10^{-42}; **Figure 5.6B**) across all markers, suggesting that altered methylation at the majority of CpG sites associated with BMI is a consequence of BMI.

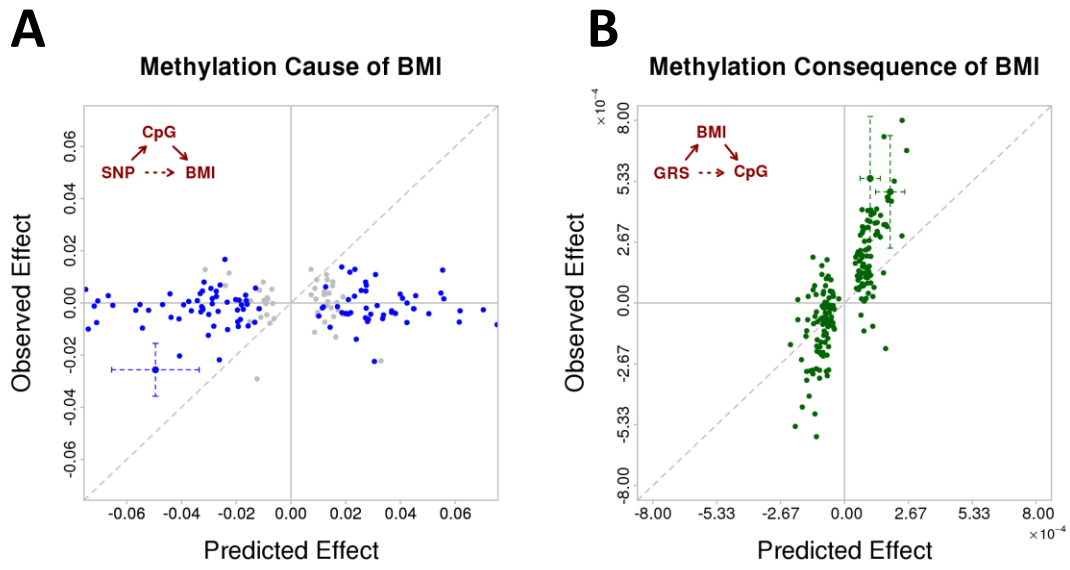


Figure 5.7: Mendelian Randomisation type analysis indentifying CpGs that are cause or consequence of BMI. A: Predicted and observed effect sizes for SNP vs BMI associations corresponding to the "causal" triangle (insert top left corner). Grey points represent CpGs not significantly associated with a SNP; blue points represent CpGs significantly associated with a SNP; for a single CpG (cg26663590) the associated SNP is also associated with BMI[86] and 95% CI error bars shown. B: Predicted and observed effect sizes for Genomic Risk Score (GRS) vs CpG associations corresponding to the "consequential" triangle (insert top left corner).

Five CpGs are significantly associated with the GRS (95% CI error bars shown). The overall correlation between observed and predicted effects suggest that the majority of CpG-sites are consequential to BMI.Two CpGs (cg00138407 and cg06500161) are significantly associated with the GRS (95% CI error bars shown). The overall correlation between observed and predicted effects suggest that the majority of CpG-sites are consequential to BMI.

5.4 Discussion

Here I have described the results of a large, prospective EWAS investigating the relationship between DNA methylation and BMI. Differential methylation of genomic DNA at 187 genetic loci associates with BMI in a meta-analysis of samples across two population groups. Moreover, the study provides proof of principle for systematic evaluation of DNA methylation as a means to find epigenetic associations with phenotypes. The lack of knowledge of the mechanisms underlying adiposity represents a major obstacle to the development of effective measures aimed at reducing the burden of obesity.

Furthermore, the 187 hits are enriched for functional genomic regions, as is shown with the enrichment for regions in open chromatin, suggesting that the hits are indeed relevant for transcriptionally active regions. The fact, however, that the hits also lie outside the expected regions of inter-individual variability that the 450k array targets, i.e. outside CpG shores, shelves and islands and inside gene bodies shows that there is still need for more knowledge about the epigenome for future array development.

The results also show that the CpG sites associating with BMI are located in potentially relevant pathways, such as insulin signalling and lipid metabolism, as well as trans-membrane transport of molecules. Further research is necessary to uncover how DNA methylation perturbs these pathways. The association of the replicated CpG sites with clinical phenotypes can potentially provide a link of BMI with these metabolic disturbances.

Causal analysis by combination of the epigenetic data with genetic data suggested that the association of the CpG sites is mainly driven by the BMI itself. This shows that the candidate genes and pathways identified lie downstream of the effect of adiposity. However, these CpGs give an indication of how adiposity might differentially affect individuals if they indeed mediate part of the association of BMI with other clinical outcomes. Further, formal Mendelian Randomisation experiments with larger samples sizes need to be carried out to characterise this relationship in more detail.

Overall, this chapter shows that differential methylation of genomic DNA from peripheral blood is strongly associated with BMI in both Indian Asians and Europeans. The findings can provide new insights into the mechanisms underlying BMI, and the basis for development of new strategies to tackle the emerging global epidemic of obesity.

Chapter 6 Enrichment for Haplotype-specific Methylation at Genetic Loci for T2D and Obesity

6.1 Introduction

In the previous chapter I outlined an EWAS and found that methylation levels at a large number of CpGs associate with BMI and related clinical outcomes. My work showed, in concordance with other observations[211, 213], that small differences in methylation can associate with risk for common diseases. However, the underlying sources of epigenetic variability still need to be explained. In this chapter, I will investigate a more general hypothesis that variation in methylation may associate with genetics through a concerted haplotype effect. I will then investigate whether there is evidence for this effect driving known associations between SNPs and methylation. Thus, I will test for enrichment of these effects at known loci for T2D and obesity.

HSM is defined as methylation that occurs via SNPs that are co-inherited on a haplotype (**Figure 6.1)**.[192] In the mechanism hypothesised[181], combined action of CpG-SNPs "in phase" (CpG creating/abrogating, **Figure 6.1**) can lead to methylation changes across a whole haplotype. From various studies, we know that methylation at neighbouring CpG sites correlates[163, 179, 306, 307], especially for those sites in the same regulatory feature.[179] While this "methylation LD" is a consequence of regulatory mechanisms affecting the same feature, it has been shown that DNA methyltransferases (DNMT) become more active in the presence of neighbouring methylated sites[308], showing that methylation can cause methylation in neighbouring regions. It then becomes plausible that an in-phase deletion of possible methylation sites can reduce methylation on a whole haplotype.

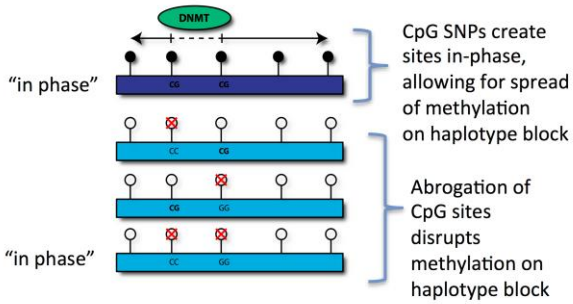


Figure 6.1: Hypothesised mechanism for the HSM effect. Methylation at neighbouring CpG sites can correlate, which may be mediated by DNA methyltransferases (DNMTs) that recognise existing methylated CpG sites. If SNPs exist on multiple CpG sites in the same haplotype region, haplotypes with a different CpG count and thus methylation potential exist. Open circles: Methylated CpG, Closed circles: Unmethylated CpG.

With respect to disease-associated DNA variants, these effects can be considered as an interplay between genetics and epigenetics, where variants in the primary DNA sequence that were previously implicated in disease risk exert their biological effects via an epigenetic pathway. This means that a subset of GWAS loci could exert effects that are mediated by changes in DNA methylation, which in turn then can have downstream effects. The downstream effects could come into effect via an association of methylation with gene expression, which in the classical model occurs by preventing binding of methyl-binding domain factors or destabilisation of the nucleosomes[309], or transcriptional splicing, as DNA methylation has been implicated in the distinction between exons and introns.[310] Moreover, DNA methylation within gene bodies can play a role in modulating transcription elongation efficiency via alternative promoters located in gene bodies[311] or by regulating non-coding transcripts related to the genes.[312] In addition, the effect of methylation on recombination rate can be another mechanism of action for genetic/epigenetic interplay.[313] The HSM effect thus provides a mechanism of action for disease associated SNPs.

A previous study showed that there is a haplotype-specific effect of methylation at *FTO*, where BMI-associated SNPs also correlated with CpG methylation levels at

nearby sites (within the first intron of the gene).[181] The study was carried out in whole blood from 60 female individuals and examined genotype-epigenotype interactions in the context of T2D, focussing on known genomic regions of susceptibility.[85] The subsets were stratified according to disease susceptibility haplotype using previously identified risk loci. Of these, significantly increased DNA methylation on the *FTO* obesity susceptibility haplotype, tagged by the rs8050136 risk allele A[104], was found. An analysis of the sequence revealed that the methylation difference was driven by the co-ordinated phase of CpG-creating SNPs across the risk haplotype. These results were then confirmed by pyrosequencing. The Bell et al. study also demonstrated that integration of GWAS SNP and epigenomic DNA methylation data can identify potential novel genotype-epigenotype interactions within disease-associated loci, thus providing a novel route to interpret common complex disease SNPs. Furthermore, as the region overlaps with the histone H3K4me1 mark, the HSM effect may further modulate enhancer activity.[181]

While the SNP association with methylation at the *FTO* locus has been replicated by several studies[181, 193], establishing whether the methylation is on the causal pathway between the risk SNP and T2D[181] has been controversial, as neither the Bell at al. [181] nor the Toperoff et al.[193] study found an association of methylation with T2D status. If it is, the *FTO* locus could represent an example of HSM.[194]

A further study by Dayeh et al.[201] recently investigated methylation at 40 known T2D loci for the presence of SNPs within CpG sites in a candidate gene study in human pancreatic islets. Of the 40 SNPs, there were 19 within a CpG site. In the study all 19 of these associated with methylation of nearby CpG sites and 91 additional CpG-SNPs in LD were also associated, which could indicate a haplotype-specific effect. Another study by the same group found that in the NADH dehydrogenase 1 beta subcomplex 6 (*NDUFB6*) gene a SNP introduces a CpG site, which then associated with increased methylation.[314]

SNPs in CpG sites are very common variations in the human genome and are already known to be likely candidates for human diseases. The nature of the methylated cytosine means that if this DNA base undergoes hydrolytic deamination, it does not become a uracil, which is foreign to DNA and specifically replaced, but a thymine instead.[315] This leads to a higher mutation rate at CpG sites, and thus more likely sites for genetic diseases.[315] Studies have suggested that 30-35% of SNP variants in the human genome are due to methylation at CpG sites.[313, 316] Additionally, an early study of allele-specific methylation found that 38-88% of the SNPs associated with methylation are located within a CpG site[180] and a link between meQTLs and SNPs in CpG sites was also suggested by other studies.[307]

Another recent study took these ideas forward and found that 80% of genetic variants at CpG sites are also meQTLs, and also have strong enrichment in the strongest meQTL signals.[317] While they find methylation effects clustering in very short regions of 10 kb, no direct link to the hypothesis of a haplotype-specific effect was shown. The largest study of HapMap cell lines[230] which overlap with the 1000 Genomes samples has been carried out by Moen et al. This provides a readily available database of both methylation and phased genotype data useful for haplotype-specific analyses.

6.1.1 Aims

Based on these recent reports, I asked the question if there is in enrichment for HSM effects in genetic loci for T2D and obesity loci. Specific aims included:

1) To find whether there is an enrichment of CpG creating/abrogating SNPs on haplotype blocks (LD blocks) of known T2D SNPs
2) To investigate if the number of CpGs on a haplotype associates with methylation across the same haplotype
3) To investigate if meQTLs found by previous studies are enriched for haplotype-specific methylation

6.2 Methods

6.2.1 T2D and obesity loci

Using a literature search, 85 robustly associated T2D- SNPs and 118 obesity (BMI, childhood obesity, adolescent obesity, class 1/2 obesity, extreme obesity, BMI, WHRadjBMI, VAT, SAT, pericardial fat, fat%, CT, NAFLD) SNPs were recorded. The individual SNPs and loci are given in **Appendix Table 4.**

6.2.2 Calculating haplotype blocks

The 1000 Genomes Phase 1v3 phased genotypes were downloaded and subset to include only CEU samples.[239] Haplotype blocks were calculated with the D' confidence interval method[318] using Haploview in regions of 500kb around each SNP. This yields haplotype block positions, SNPs included, and haplotype frequencies for each haplotype allele.

6.2.3 CpG Score

Once the haplotypes were obtained, at each block the haplotypes were reduced to only contain SNPs within CpG sites. For each haplotype, the sum of intact CpG sites was recorded. To match the average count of CpG sites to each T2D SNP risk allele, the sum of the CpG sites per haplotype containing the SNP allele was recorded in a weighted averaged weighted by haplotype frequencies. This process is outlined in **Figure 6.2**. The difference between these two averages was then used as a score for the in-phase CpG creating/abrogating potential of the SNPs.

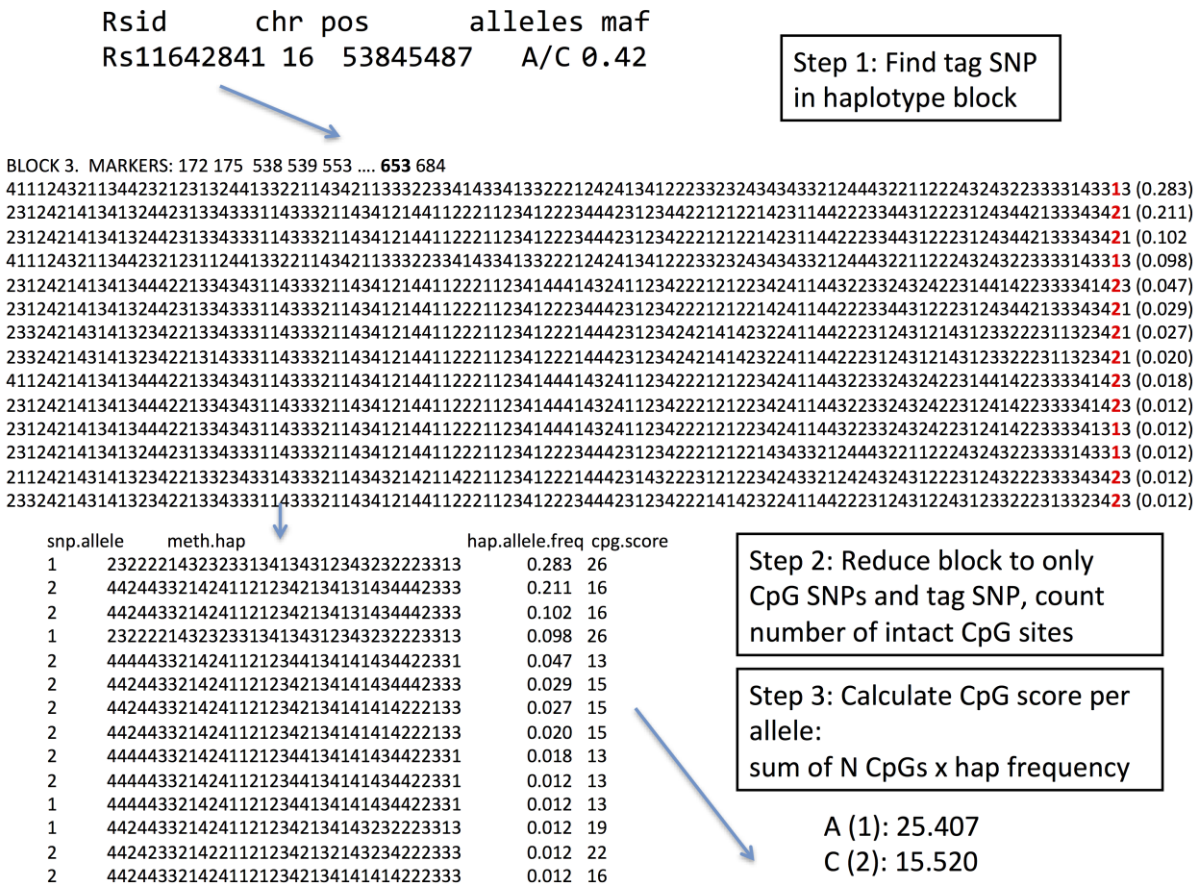


Figure 6.2: Illustration of calculation of the HSM CpG Score based on the *FTO* locus. 1) The SNP provided from the publications is looked up in the phased 1000 Genomes data. If the SNP is on a haplotype block, the individual haplotypes and haplotype frequencies are retrieved. 2) The haplotype blocks are then reduced to only the SNPs within CpG sites 3) From these, per allele a CpG score is calculated as the sum of CpGs x haplotype frequency.

6.2.4 Permutation Testing

To test haplotype scores and random permuted loci, I devised a strategy to streamline the computationally intensive process outlined above. First, I sampled a number of common SNPs (MAF > 5%) from the 1000 Genomes reference, inferred their haplotypes and calculated the CpG Scores as outlined above. This was completed for a pool of 43,000 SNPs (0.70% of the SNPs in the panel). Secondly, to carry out permutation testing, I matched SNPs to the T2D and obesity SNPs by selecting SNPs from the pool to match 1) MAF ±5% and 2) the same number of CpG SNPs in the haplotype. This should account for both CpG content and haplotype length. I then used the distribution of CpG difference scores amongst randomly sampled SNPs to establish a null-hypothesis and calculated an empirical p-value for each observed disease SNP as the proportion of the CpG score differences that were as high or higher than the observed CpG score difference.

6.2.5 Global Analysis of haplotypes

To test whether the CpG score indeed predicts methylation, I downloaded the 450k data for 32 European and 51 African HapMap samples[230], and subset them to include only the European samples. To test for association of methylation, I calculated a CpG score for each individual as the number of SNPs with intact CpG sites at each haplotype, as an average of the scores on both haplotypes, creating a single score for each individual at each haplotype. The number of CpG SNPs was then directly fitted as a predictor against methylation level using a linear model. This was done around a subset of methylation probes representing ~10% of the array.

6.2.6 Enrichment analysis of previously observed meQTL

To test if CpG Score explains part of the previously published meQTL hits[178, 179], I carried out an analysis similar to the T2D and obesity enrichment. The list of

published meQTL was then taken as an input in place of the list of T2D and obesity SNPs to test whether the meQTL hits are enriched for associations.

6.3 Results

6.3.1 HSM effect of T2D and Obesity SNPs

I tested haplotypes of 85 SNPs currently known to be robustly associated with T2D and 118 SNPs associated with Obesity traits (**Appendix Table 4**) for an enrichment of CpG creating or abrogating SNPs. Using this method, I was able to score 81/85 T2D SNPs and 106/118 obesity SNPs. The remaining SNPs could not be assigned to a haplotype block.

To test whether the observed CpG score difference is higher than for any other SNP in a similar genomic context of the disease SNPs, I first randomly sampled SNPs from the autosomal 1000 Genomes Data and calculated the CpG score difference. This selection procedure controls both for haplotype length and CpG content.

Using this approach, I find 9 T2D (**Table 6.1**) and 23 Obesity SNPs (**Table 6.2**) that are enriched at $p < 0.10$ for CpG creating/abrogating SNPs. Amongst these is *FTO*, which is associated with T2D via BMI, a locus which has been previously been shown to have a HSM effect[179, 193, 194], which suggests that our method can identify true positives. Other loci of interest are *POMC*, where CpG methylation has been shown to associate with childhood obesity and *BDNF*, a transcription factor, where promoter methylation has been implicated in the regulation of food intake.[319]

Table 6.1: T2D loci with an HSM effect p < 0.1.

Chr	Position	rs ID	AF	A1 CpG Count	A2 CpG Count	total CpG SNPs on hap block	N random SNPs	P-value	Adjusted p-value	Locus
2	43732823	rs7578597	0.09	23.00	10.62	26	440	<0.002	0.000	*THADA*
16	53845487	rs11642841	0.42	25.41	15.52	35	247	0.020	0.719	*FTO*
1	120517959	rs10923931	0.11	9.12	4.00	12	1223	0.033	0.719	*NOTCH2*
10	80942631	rs12571751	0.46	16.58	23.82	26	334	0.036	0.719	*ZMIZ1*
3	123065778	rs11708067	0.18	15.52	22.62	28	422	0.055	0.786	*ADCY5*
13	80717156	rs1359790	0.27	1.00	3.00	3	1231	0.066	0.786	*SPRY2*
15	90374257	rs2028299	0.73	11.48	17.80	24	420	0.069	0.786	*AP3S2*
2	60568745	rs243088	0.47	1.00	3.72	5	769	0.079	0.786	*BCL11A*
1	214159256	rs340874	0.54	5.00	2.33	5	803	0.088	0.786	*PROX1*

Chr: Chromosome, hap: haplotype, AF: allele frequency, A1: allele 1, A2: allele 2, N random SNPs: Number of matching SNPs in permutation test, N random SNPs: number of random SNPs matched in sample, p-value: Permutation based p-value, Adjusted p-value: FDR-adjusted p-values

Table 6.2: Obesity loci with an HSM effect p < 0.1.

Chr	Position	rs ID	AF	A1 CpG Count	A2 CpG Count	total CpG SNPs on hap block	N random SNPs	p-value	Adjusted p-value	Locus
4	103188709	rs13107325	0.08	49.43	39.83	82	16	<0.063	<0.063	*SLC39A8*
16	53813367	rs17817449	0.42	26.00	15.54	35	247	0.012	0.135	*FTO*
5	95850250	rs261967	0.42	7.32	15.66	19	470	0.013	0.135	*PCSK1*
16	53825488	rs9941349	0.43	15.39	25.88	35	232	0.013	0.135	*FTO*
1	177855517	rs516636	0.2	13.00	6.75	14	776	0.015	0.135	*SEC16B*
2	25158008	rs713586	0.46	7.89	2.00	11	602	0.017	0.135	*POMC-RBJ*
16	53842908	rs12149832	0.43	25.79	15.46	35	232	0.022	0.163	*FTO*
11	27725986	rs10767664	0.76	7.08	13.00	15	588	0.029	0.192	*BDNF*
11	27728539	rs2030323	0.76	13.00	7.08	15	588	0.029	0.192	*BDNF*
1	72812440	rs2815752	0.63	3.96	8.53	12	604	0.051	0.307	*NEGR1*
5	55859952	rs11743303	0.18	3.67	0.91	4	1556	0.052	0.307	*MAP3K1*
16	53800954	rs1421085	0.44	3.12	7.34	10	802	0.059	0.311	*FTO*
16	53803574	rs1558902	0.44	3.12	7.34	10	802	0.059	0.311	*FTO*
22	44324727	rs738409	0.22	2.00	0.24	2	1360	0.087	0.438	*PNPLA3*
22	29451671	rs4823006	0.42	1.31	3.91	5	1007	0.092	0.445	*KREMEN1*
2	12964497	rs10198628	0.55	26.24	19.70	36	193	0.098	0.453	*TRIB2*

Chr: Chromosome, hap: haplotype, AF: allele frequency, A1: allele 1, A2: allele 2, N random SNPs: Number of matching SNPs in permutation test, N random SNPs: number of random SNPs matched in sample, p-value: Permutation based p-value, Adjusted p-value: FDR-adjusted p-values

6.3.2 Global Analysis of haplotypes

I then set out to detect whether the number of CpG SNPs on a haplotype block influences methylation on a nearby CpG site. To achieve this, a similar procedure was carried out for 10% of the CpG positions the 450k array. CpG sites were only investigated for association of the haplotype block directly overlapped with the CpG site in question, and p-values of association calculate as described above.

Overall, I find a strong enrichment for low p-values (**Figure 6.3**), indicating that there is indeed a global effect of the number of CpGs on methylation level. Scatter plots showing the relationship between haplotype CpG counts and methylation at the top 4 associated loci are shown in **Figure 6.4**. HSM can be in effect at haplotypes with a small number of CpGs, as well as large numbers. In most cases, methylation clustered into three distinct clusters, potentially corresponding to a single SNP genotype.

Table 6.3: Global enrichment of association of CpG score with methylation. A linear model of CpG count against methylation beta value was fitted for each CpG lying within a SNP haplotype and p-values of association were recorded. The QQ plot shows the enrichment of associations.

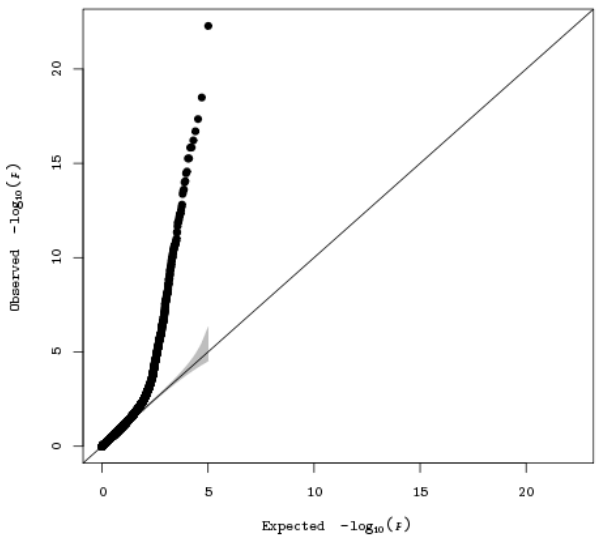
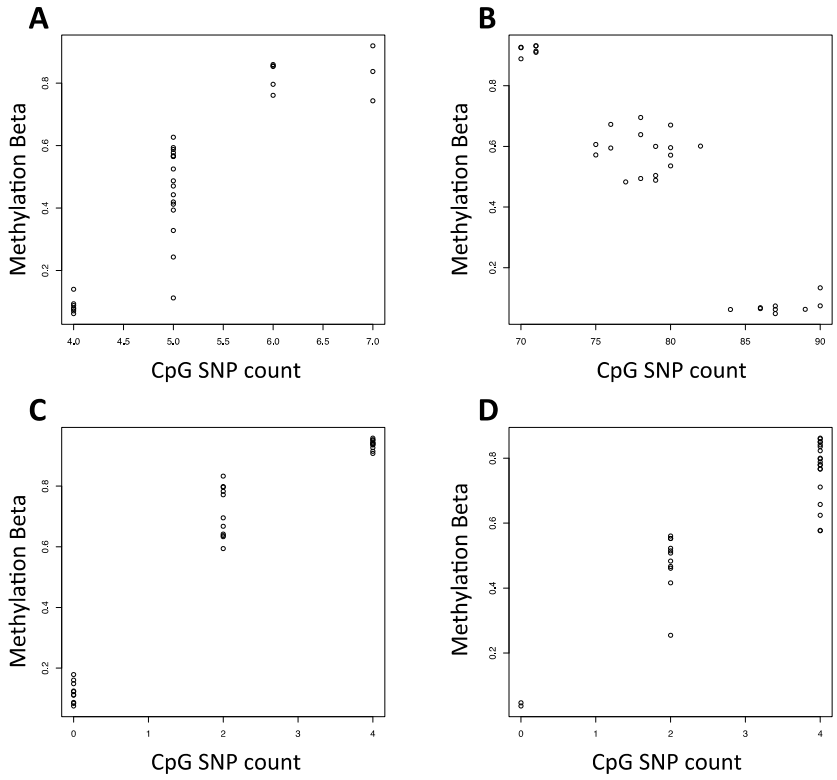


Table 6.4: Top four associations of CpG count across a haplotype with methylation at an overlapping CpG site. A linear model of CpG count against methylation beta value was fitted for each CpG lying within a SNP haplotype and p-values of association were recorded. The scatterplots directly show the relationship between CpG count and methylation beta at the top four loci. A: cg02890259, B: cg06961873, C: cg08146708, D: cg08950364



I next asked the question whether the observed HSM effect can explain a number of meQTLs found by previous studies. Of the three studies I investigated (Bell et al. HapMap 27k data in LCLs[179], Gibbs et al. 27k data in various brain tissues[178], and the more recent Moen et al. study of HapMap LCLs[230]) none showed enrichment for HSM effects.

6.4 Discussion

My work showed for the first time that in-phase creation or abrogation of CpG sites by genetic variants can associate with methylation levels at a nearby CpG sites. Moreover, I found that the known T2D and obesity SNPs are enriched for variants that segregate on haplotypes with a strongly varying number of intact CpG sites

between each SNP allele. This provides a potential mechanism of action for the many non-coding variants found amongst hits for these common diseases.[86, 128]

While I found that the CpG count on a haplotype associates with methylation of CpG sites on the haplotype directly throughout the genome, I did not see any enrichment for previously published meQTLs. This could be due to the restricted nature of my search (focussing on very close-by CpG SNP associations), or an inaccuracy in the way the CpG score metric is simply calculated as the difference between the two SNP alleles. The fact that the differential count at *FTO* was replicated, however, was a first confirmation that the method can indeed point to HSM effects.

The number of CpGs on the haplotype is measured with the best possible knowledge, using fully phased 1000 Genomes data, the aggregation of haplotypes to one SNP allele can probably be improved from simply weighting by haplotype frequency. It can also be imagined that SNPs within important regulatory regions play a more important role, than SNPs far away from the CpG site. A better weighting procedure could benefit the CpG-SNP methylation prediction. Also, in the current linear model the effect is considered to be additive. Instead of simply calculating the difference in CpG count per risk SNP allele, a fold-change could be imaginable, or fixed thresholds acting as an “on-off switch” for methylation across the whole haplotype.

However, despite the weaknesses in the original model, there is a strong enrichment of HSM already visible. With more knowledge of genetic drivers of associations, one could imagine that predicting methylation status would become possible with knowledge of nearby genetics, similar to studies done for eQTLs.[320, 321] These methylation values could then still be modulated by environmental effects, but as genetic effects can be strong, these CpG SNPs could then serve as a good basis to predict the approximate methylation level.

Possible improvements of the pipeline could also include more accurate matching of the sampled SNPs. While the absolute number of CpG SNPs is an approximate measure for both the length of the haplotype and its CG, as well as CpG content, the latter variables can be recorded directly to improve matching.

A further problem is that the haplotype blocks calculated are not always clearly delimited. As neighbouring haplotype blocks also have a haplotype LD, one could imagine the HSM effect to act across two or more haplotypes. One way to take this into account in future analyses would be by broadening regions and taking into calculate the LD between neighbouring haplotype blocks.

Future work could investigate the HSM effect in whole-genome bisulphite sequencing data. The main benefit of these data is that the methylation can be measured both specific to a strand and chromosome if nearby SNPs are also recorded. Rather than using 450k data, which would report an average of the two chromosomes, this would then directly give a link between the relationship of genetic variation with methylation within the same DNA molecule.

With respect to the enrichment for T2D and obesity loci, future work could be carried out on a large case/control study where these the fully phased genotypes are available, which could then lead to a full investigation of CpG score differences at each recorded haplotype in order to gain a better understanding of how genetic variants interact. This could potentially give an explanation for the biological effect of observed signals. These studies can further be supported by data sets larger public efforts, such as ENCODE[131, 132] and the IHEC[322], once more population level data has been generated.

Chapter 7 General Discussion

The work in this thesis aimed to explore the relationship of epigenetics with T2D and obesity and followed five specific aims:

- To find genetic variants associated with DNA methylation in adipose tissue
- To find factors confounding EWAS
- To evaluate and develop methods for EWAS
- To carry out an EWAS for BMI
- To investigate haplotype-specific effects at known T2D and obesity loci using bioinformatics.

My work outlined in Chapter 3 established that there are a number of genomic regions in adipose tissue where methylation is controlled by genetics. I combined methylation data from abdominal SAT samples with genotypes and mRNA expression levels all collected from 38 unrelated individuals. Here, I identified 149 significant meQTLs of which five replicated in 181 independent individuals. While none of the meQTLs were also significant eQTLs in my data, the expression levels of two mRNA transcripts associated with *cis*-methylation status.

My work presented evidence that meQTLs, which have been established in other tissues[178, 181], are present in fat. As time has passed, the DMH technology used in this project[234]has been superseded by the Illumina BeadChip, but the replication in the independent cohort using the Illumina 27K array provided validation for the DMH methodology. As there is relatively little overlap between these array technologies, it is still not clear whether the meQTL in other tissues (LCL[179] and brain tissues[178]) did not replicate for technical reasons or power issues, or whether the meQTL are indeed specific to adipose tissue. Since the meQTL findings were published, another study of meQTLs in adipose tissue, skin and blood was carried out by the MuTHER consortium.[232] The study, which investigated 648 adipose tissue samples and had

thus increased power, found that a very large number of the CpG sites on the array (28%), show association with a SNP nearby. Moreover, they found that 6% of these loci overlap with eQTLs. On the other hand, in the work presented here I find no overlap between meQTL and eQTL, which would be consistent with the lower power and generally low overlap observed in the MuTHER study. To fully understand the genetic contribution of SNPs on methylation, more power and replication studies will be necessary.

The Pilot Study in Chapter 4 investigated global effects on methylation in a longitudinal, nested case control study. I was able to find that there were wide-ranging systematic effects on methylation with the onset of T2D, but no differences between subjects who were or were not on treatment. In line with previous reports studying the variation of methylation with aging[163, 179], I found that methylation data, as measured with the 450k, were highly different between the two time points at which methylation was measured. Moreover, the differences between beta values were significantly different between the controls and the T2D groups. I further found that WBC subsets were a major determinant of the methylation patterns observed.

The associations of methylation with age have been investigated by other groups as well since the Pilot Study was undertaken, and our results are largely consistent.[252, 264, 284, 285]. One study even claim that the epigenetic profile can be used as an "epigenetic clock" to determine the biological age of a person.[285] After I carried out the Pilot study a method was published for inferring WBC subsets in methylation data[270]. This somewhat relaxed the requirement to have cell measurements, but the decision was still useful, as the measured data could be used to validate the estimated cell proportions. As at the time no method was available to calculate these retrospectively, it was decided to also only include samples that have cell proportions measured. Similarly, confounding by the use of retrospective cases is not surprising, and the use of incident cases has previously been suggested.[192, 323]

To avoid confounding in the main study, we decided to not include any prevalent cases, but rather incident cases, which were NGT at the time the sample was taken.

With the complete EpiMigrant data set, I then resolved issues in the QC and normalisation for the development of a pipeline (Chapter 4). First, I evaluated the routinely used detection p-value threshold used for QC by using three metrics: call rate on chromosome Y, outlier rate on autosomes and the reproducibility between duplicates. Next, I assessed the performance of various published normalisation methods, and found that a quantile procedure performs best. Permutation testing revealed that the correlation structure of methylation data is wide ranging, as a broad prediction interval showed inflation or deflation in a large number of permutations. The correlation structure was mainly driven by technical factors and inflation is lowered if these confounding factors (WBC proportions and control probe PCs) are included.

My work highlighted the need for quality metrics to assess 450k data quality and complements methods developed in parallel in the field. Pidsley et al.[324] used three different metrics to asses normalisation methods: i) behaviour of the beta value at imprinted genes (expected to be ~0.5), ii) clustering of the intensities of the 65 probes that measure SNP genotypes and iii) the difference between signals on chromosome X markers between males and females.[324] Using their independent methods on a set of 696 samples, they also find that a quantile normalisation procedure of intensities (*dasen* in the *watermelon* package)[324] performs best in terms of reproducibility. Another set of observations was published by Wu et al.[325] that assessed four normalisation methods (SWAN, Illumina normalisation, BMIQ and the Tost pipeline) using 14 samples measured in duplicate. The performance of these methods in their analysis is similar to the performance in my analysis of 36 samples measured in duplicate, but quantile normalisation of intensities was unfortunately not considered in the Wu et al. analysis.[325]

The importance of adjusting for estimated WBC subsets was also highlighted in a recent publication by Jaffe et al.[326], which found that many of the observed differences for age were simply due to differences in WBC counts.[326] While other "complete" pipelines have been developed in parallel[219, 281], many of these do not take the batch effects and correlation structure of the data into account. Moreover, as the methods developed in Chapter 4 have been optimised for the use on computing clusters, the single-marker tests are fast and can deal with much larger datasets than those on desktop machines. While in other settings, such as those of regional analyses, or analyses of smaller datasets with a more complicated structure, a custom pipeline would be more appropriate.

I used the pipeline outlined in Chapter 4 for my work described in Chapter 5. Here, I investigated whether methylation levels associate with BMI and carried out a meta-analysis of these association results in independent samples provided by collaborators.[86, 87] I found 187 CpGs at which methylation levels are consistently and significant associated with BMI. Genes close to the differentially methylated CpGs are enriched for pathways in insulin signalling and metabolism, but also co-cluster with histone marks representing open chromatin and/or promoter/enhancer sites. Furthermore, methylation at the 187 replicated loci also associated with other clinical outcomes, and associations even persisted after adjusting for BMI. As a first preliminary test to see if the CpG sites explain the variability in these phenotypes, the percentage of the association of BMI with the individual traits was calculated to retrieve the proportion of this indirect effect, suggesting that association of BMI with these phenotypes was in part mediated methylation at the CpG sites.

To my knowledge, my work provides the largest EWAS of BMI to date and superseded a previous effort by Dick et al. finding an association of DNA methylation with BMI in whole blood at the *HIF3A* locus.[213] The fact that the *HIF3A* locus[213] previously associated with BMI did not replicate could have multiple reasons: For

example, by confounding in the previous study from using data from different centres, myocardial infarction cases and controls or different normalisation methods.[213]

My results in the EWAS work also give some insights about future array design. I find that the associations were not in the CpG islands, shelves or shores, but in “open sea” locations, which were not the focus of the array development.[191] Also, many of the associations lie in gene bodies, which have been implicated more recently to show a complex relationship with expression.[311] The result that the associated CpG sites are depleted for CpG Islands was not surprising, given the previous work showing that variability happens in other regions.

The next question the results raise is whether the observed signals are due to reverse causation, or subtle effects that have not been accounted for. In work published by Dick et al.[213], the issue of causality was addressed by integrating SNP data. Genetic variants were found that associate with methylation levels at the *HIF3A* CpG sites, but not BMI, which consistently suggests that differential methylation associated with BMI is a consequence of BMI rather than the cause.

Another EWAS on metabolic traits has been carried out in parallel to my work. In an EWAS of 544 discovery and 293 replication samples for fasting glucose fasting insulin and HOMA-IR, Hidalgo et al.[202] found a significant association of methylation at *ABCG1* CpG sites with fasting insulin and HOMA-IR. The CpG site found to be associated with the traits, cg06500161[202], also associated significantly with BMI in Chapter 5. In an analysis of SNP data around the CpG site, Hidalgo et al. showed that SNPs associated with differential methylation at the CpG site only marginally associated with the two metabolic traits[202], again raising the question of causality.

Moreover, the question whether methylation in blood is a proxy for methylation in other tissues has been addressed when checking for methylation levels across tissues. In this study, the methylation in blood was similar to omental and

subcutaneous fat, as well as skin. This showed that methylation in blood can be a useful proxy for methylation in other tissues. Similarly, Dick et al showed a consistent association of the differentially methylated CpG sites in adipose tissue.[213]

In the final chapter, Chapter 6, I investigated whether any of the known T2D genetic loci are enriched for regions of HSM, which represents an integration of both genetic and epigenetic effects. Using a publically available 450k methylation data set that matches part of the samples within the 1000 Genomes, I found that at many haplotype blocks, the number of intact CpG sites correlates with methylation levels on the same haplotype. Further, I found a global enrichment and detected that haplotypes at the *FTO* locus have a differential CpG count with this novel method.

My work here work ties in with previous reports that T2D loci are enriched for CpG-changing SNPs[327], and that CpG SNPs can drive the association of meQTLs.[317] I find no evidence of a general enrichment for HSM effects amongst at set of published meQTL hits.[178, 179, 230] This could mean that the HSM effect does contribute towards associations of methylation with genetics, but only represents a pathway active at a certain subset of meQTLs. The next steps could involve larger meQTL studies where confidently phased haplotypes are available, for which the scripts that I generated for this study could be used as they rely on input from commonly used genotype and methylation files. If in the future whole-genome bisulphite sequencing becomes more readily available, it becomes feasible to assign the methylation values directly to each haplotype, which can yield much more power in the above-mentioned analyses.

Overall, the EWAS presented in this work has followed a similar trajectory as GWAS, moving away from candidate gene studies to large collections of cohorts doing global genome wide analysis, and on to subsequent meta-analyses of multiple studies to increase sample sizes even further. My results also show the usefulness of collecting methylation in blood, as the associated CpGs seem to centre on genes involved in

metabolism, and correlate well with methylation patterns in other tissues. In the future, similar studies with equal sample size can be carried out in other, more relevant tissues to solve these questions, but there are still limitations by the tissue collection process.

All the analyses that I have carried out have been on tissues, which are heterogeneous as they consist of an array of different cell types. This heterogeneity can be overcome by collecting methylation values for flow-sorted reference panels for whole blood samples[270] and more and more reference panels will become available.[328] Such data will further help untangle spurious associations due to cell types when carrying out tissue specific EWAS. As the heterogeneous cell populations can also generate more noise in the data, this will mean that tissue experiments also need a higher sample size than single cell analyses. When conducting single cell experiments, one benefit is that these will be more powerful on a per-sample basis, but suffer from generally low throughput in the collection or processing. A further criticism can be that in the single-cell case the cells are taken out of their biological context and are more prone to batch effects created in the experimental process.

Other improvements in the EWAS field that will likely take place over the next few years will be the advancements in assay technology. While the lessons learnt from the 450k array, such as the remaining low coverage and bias have become apparent in this work, there will certainly be a successor generation representing a step-up that can be compared with that between the 27k chip and the 450k. In parallel, as sequencing will drop in price, population-wide whole genome bisulphite sequencing and more targeted enrichment analyses such as reduced representation bisulphite sequencing will become cheaper.

To establish causality in a formal way, the concept of Mendelian Randomisation (MR)[293, 329] is becoming more and more useful. As being advocated by previous works[293], this concept makes use of the fact that germline genetic variants cannot normally be changed during lifetime. An MR analysis can then include the association of the genetic variants (or genetic risk score of combined variants) with the phenotype of interest, and the intermediary, in the case of EWAS the methylation. MR takes into account that the genotypes would be "randomised" between cases and controls, and that if methylation is on the causal pathway, any genetic variant associated with methylation would then necessarily also be associated with the phenotypic outcome, in this case, methylation.

The data presented in this thesis have addressed central questions in the field of epigenetic studies. Following the development of the field, I worked on tissue-specific small-scale pilot studies (Chapters 3 and 4), to method development (Chapter 4), all the way to an example of an EWAS and subsequent meta-analysis (Chapter 5) to novel methods in bioinformatics (Chapter 6) for exploring the effect by which genetic variants act on methylation.

Appendices

Appendix Tables

Appendix Table 1: Associated meQTLs

DMH ID	Chr	Probe Start (b37)	Probe End (b37)	P-value	SNP rs ID
EpiGII1027953Q	3	176822139	176822281	8.38E-12	rs880210
EpiGII1035732Q	3	179040338	179040601	1.16E-06	rs10513760
EpiGII1052897Q	3	184052376	184052634	2.71E-06	rs10937181
EpiGII1083066Q	3	192778112	192778492	4.93E-13	3-192773535
EpiGII119081Q	1	228075665	228075925	9.99E-08	rs6690181
EpiGII1278944Q	5	63700343	63700504	2.16E-06	5-63267358
EpiGII1423698Q	5	78281698	78281893	6.11E-07	rs62377912
EpiGII1510592Q	6	74230015	74230448	2.14E-07	rs35592974
EpiGII1619588Q	6	165729959	165730386	1.66E-06	rs617910
EpiGII1622451Q	6	166582675	166582968	1.47E-06	rs6928596
EpiGII1625292Q	6	167450032	167450170	2.85E-06	rs2236312
EpiGII1689143Q	6	26568897	26569208	2.72E-08	rs16891264,6-26072981,rs7749414
EpiGII1707954AG	6	32188833	32189031	5.63E-07	rs520803
EpiGII1710072AG_x	6	32820694	32820884	2.96E-06	rs28624333
EpiGII1735794Q	6	40408884	40409047	2.16E-06	6-40744965
EpiGII1737362Q	6	40860251	40860501	1.64E-06	6-40774036
EpiGII1806709Q	7	158503157	158503349	4.83E-07	rs11971680
EpiGII1921930Q	7	21945979	21946208	2.18E-09	7-21946177
EpiGII1997158Q	7	43495893	43496105	2.82E-06	7-43434552
EpiGII2150420Q	7	100478877	100479203	1.86E-06	7-100937964
EpiGII2154525AG	7	101851903	101852365	3.31E-06	rs62463713
EpiGII223970Q	1	203148336	203148707	7.89E-07	rs10399805
EpiGII2264054Q	7	133812100	133812441	6.01E-07	rs13243553
EpiGII2424610Q	8	119611743	119611899	2.19E-06	rs4336638
EpiGII2504294Q	8	54851968	54852304	1.56E-06	rs28688342
EpiGII2659339Q	9	7799281	7799539	1.85E-06	9-7455831
EpiGII2681860Q	9	14349378	14349733	2.85E-06	9-14785981
EpiGII2760967Q	9	37161648	37161807	6.12E-08	rs17408134
EpiGII2771867Q	9	99259319	99259718	2.13E-07	rs10761013,rs10739886,rs7037607
EpiGII2825376Q	9	114695104	114695251	6.97E-08	rs7872317,rs10759560

DMH ID	Chr	Probe Start (b37)	Probe End (b37)	P-value	SNP rs ID
EpiGII2856776Q	9	123622356	123622808	2.14E-07	rs1803347
EpiGII2868506Q	9	127116979	127117345	4.29E-07	rs10986284
EpiGII2882474Q	9	131413221	131413561	1.66E-16	rs10819406
EpiGII3042175Q	10	27001322	27001477	6.56E-09	rs11015202,rs11015203
EpiGII3054918Q	10	30721783	30721962	2.23E-06	rs10826797
EpiGII3082929Q	10	128810210	128810641	1.87E-07	rs9418754
EpiGII3276030Q	11	44219361	44219639	1.13E-08	rs7946749
EpiGII3398764Q	12	124429995	124430301	2.97E-06	rs10744173
EpiGII3403632Q	12	125976178	125976447	3.18E-06	12-125812398
EpiGII3420626Q	12	130920696	130920884	1.92E-06	rs34584432
EpiGII3447162AG	12	7066611	7066922	1.48E-06	rs11064295
EpiGII3448408Q	12	109675029	109675290	1.43E-12	rs3742027
EpiGII3449118Q	12	109885186	109885349	9.45E-09	12-109854511
EpiGII3619510Q	15	41625250	41625595	2.83E-06	15-41323360
EpiGII3752536Q	15	80537394	80537590	6.41E-09	15-80341682
EpiGII3781050Q	15	91016890	91017186	1.93E-06	15-91316878
EpiGII3811572Q	16	8954289	8954493	1.88E-10	rs41305755
EpiGII384317Q	2	234228968	234229357	1.56E-06	rs3828307
EpiGII3947657Q	16	71382591	71382874	3.02E-06	16-71095249
EpiGII3964139Q	16	50827428	50827594	1.35E-10	rs17222902
EpiGII3969104Q	16	89814984	89815216	7.01E-07	rs383027
EpiGII4015783Q	17	73662899	73663153	1.41E-06	17-73294294
EpiGII4030066Q	17	80708128	80708322	2.69E-06	rs9914081
EpiGII4030067Q	17	80708321	80708551	2.76E-06	17-80479582
EpiGII4128245Q	17	48166519	48166861	7.24E-07	rs61541934
EpiGII4152050Q	17	55038512	55038633	3.35E-06	rs12601451
EpiGII4232497Q	18	3360822	3361029	6.10E-08	rs1628565,rs1629295
EpiGII4266414Q	18	13241933	13242121	1.98E-07	rs7228009
EpiGII4279590Q	18	77713320	77713457	1.00E-06	rs7231668
EpiGII4369274Q	18	44087519	44087899	2.17E-07	rs11082532
EpiGII4424875Q	19	36024159	36024306	3.33E-06	rs1005948,rs8113518
EpiGII4435517Q	19	39390520	39390885	1.48E-06	rs4310986
EpiGII4451171AG	19	44171583	44171876	4.41E-08	rs344787
EpiGII4452275Q	19	44489036	44489179	4.80E-07	rs17560331
EpiGII4484467Q	19	54802668	54803094	2.81E-07	rs443501
EpiGII4507824Q	19	3975665	3975964	9.99E-07	rs41276860
EpiGII4507933Q	19	4012417	4012679	2.01E-06	rs350843
EpiGII4542582Q	19	16680992	16681330	2.20E-06	19-16869679
EpiGII4551985Q	19	19856063	19856202	1.65E-06	rs247759
EpiGII4642884Q	20	55502291	55502503	2.91E-06	rs4638857
EpiGII466643Q	2	157178044	157178306	2.24E-06	rs298288

DMH ID	Chr	Probe Start (b37)	Probe End (b37)	P-value	SNP rs ID
EpiGII4848051Q	21	45173461	45173813	4.71E-07	rs56340987
EpiGII4880881Q	22	24379215	24379420	3.01E-06	rs2739329
EpiGII4904336Q	22	31364890	31365105	1.57E-09	rs55694472
EpiGII4925937Q	22	37875303	37875658	1.43E-07	rs5750478
EpiGII4933437Q	22	40314259	40314500	3.12E-06	rs4820380
EpiGII4939537Q	22	42180723	42180893	5.57E-07	rs132792
EpiGII530184Q	2	175618003	175618266	8.91E-07	rs62175269
EpiGII5462001Q	2	20416905	20417113	2.22E-06	rs2230924
EpiGII571186Q	2	187454389	187454632	2.12E-06	2-187446311
EpiGII5814139Q	10	135175609	135175812	5.45E-08	rs3008338
EpiGII5913129Q	12	96309784	96309915	2.26E-16	rs7976481
EpiGII6020838Q	1	220393463	220393932	1.73E-06	rs10746396
EpiGII6040485Q	1	8929287	8929536	1.52E-12	rs2765507,rs2781064
EpiGII6054451Q	1	230312948	230313162	6.57E-07	rs636810
EpiGII6056256AG	1	230848290	230848518	8.30E-08	1-231110901
EpiGII6057069Q	1	231081069	231081203	6.06E-07	rs59352069
EpiGII6114193Q	2	126229629	126229881	2.60E-06	rs35304211
EpiGII6215811Q	2	35696389	35696578	1.94E-06	rs62141964
EpiGII6246307Q	2	44528169	44528407	3.44E-07	2-44516631
EpiGII6280963Q	2	54566795	54566974	5.74E-11	rs10196726
EpiGII6338546Q	2	71189758	71190042	1.46E-06	rs11466186
EpiGII6591246Q	3	27761925	27762141	2.73E-06	rs2371108
EpiGII6641328Q	3	42056200	42056461	2.95E-06	rs9876105
EpiGII6676527Q	3	52559034	52559314	2.51E-06	rs2015971
EpiGII6710050Q	3	62180451	62180974	3.16E-06	3-61909077
EpiGII6858440AG	4	57921746	57922018	2.73E-06	4-57709199
EpiGII6866201Q	5	499892	500263	3.30E-06	5-496669
EpiGII6922405Q	5	17232490	17232651	1.10E-06	5-17389052
EpiGII6974226Q_X	5	169726649	169726772	1.25E-06	5-169495516
EpiGII6995343Q	5	175985022	175985157	2.14E-06	rs2963298
EpiGII703218Q	2	225882986	225883115	3.35E-06	2-226158651
EpiGII7060808Q	8	637363	637768	2.64E-06	rs4735891
EpiGII7067482Q	8	2794191	2794368	2.84E-06	rs11782158
EpiGII7249603Q	13	42657422	42657588	2.78E-07	rs12584909
EpiGII7278174Q	13	50948524	50948776	1.45E-10	rs3095160
EpiGII7281282Q	13	51825548	51825835	2.71E-09	rs4942979
EpiGII7444628Q	16	85100149	85100449	1.59E-07	rs2934466
EpiGII7468618Q	17	78143699	78144026	2.31E-06	rs2361704
EpiGII7763234Q	6	148504339	148504841	1.79E-06	rs208697
EpiGII7792774Q	6	157083125	157083453	6.08E-07	6-156584231
EpiGII7811731Q	14	23058849	23059066	1.63E-06	rs5742883

DMH ID	Chr	Probe Start (b37)	Probe End (b37)	P-value	SNP rs ID
EpiGII791055Q	3	108125624	108125769	1.79E-06	rs840289
EpiGII7924998Q	14	55907406	55907710	2.86E-06	rs10140857
EpiGII7987921Q	14	74100455	74100981	1.43E-06	rs7158978
EpiGII8063530Q	14	96152830	96153095	9.26E-07	14-96520090
EpiGII8064943Q	14	96552916	96553157	2.87E-06	rs34019046
EpiGII8065488Q	14	96707477	96707660	2.68E-06	rs745378
EpiGII8066253Q	14	96944751	96945064	2.54E-06	rs35750540
EpiGII8081099Q	14	101349641	101349763	8.52E-07	rs11627690
EpiGII8093429Q	14	105420581	105420927	1.60E-06	rs2841268
EpiGII8236867Q	8	143125546	143125940	1.08E-07	8-143125433
EpiGII8359259Q	12	49207956	49208234	1.87E-07	12-49208045
EpiGII8360968Q	12	49735775	49736170	1.17E-06	rs7971455
EpiGII8377753Q	12	54756481	54756733	8.59E-07	rs12426667
EpiGII8469763Q	1	108735403	108735735	4.95E-07	rs1419000
EpiGII8472823Q	3	196208957	196209090	6.83E-07	3-195902418
EpiGII8511477Q	10	88282242	88282433	1.95E-06	rs2588282
EpiGII8555714Q	10	101143419	101143599	2.54E-07	rs11189951
EpiGII8572191Q	10	106035156	106035311	2.89E-06	rs11191967
EpiGII8630202Q	10	122738941	122739326	2.30E-06	rs34558886
EpiGII8646955AG	8	12953993	12954224	2.02E-06	rs17787049
EpiGII8680512Q	17	19311052	19311172	1.36E-07	rs56191443
EpiGII8688331Q	20	60085186	60085645	4.90E-07	rs6061373
EpiGII8761308Q	1	53392620	53392784	3.16E-06	rs11206009,rs61768344
EpiGII8918930Q	11	118133081	118133279	2.16E-06	11-117772389
EpiGII8954893Q	11	128615495	128615842	9.70E-07	rs2604237
EpiGII9222676Q	5	120952053	120952388	1.54E-09	rs10044156
EpiGII9342901Q	11	319335	319785	6.97E-07	rs3847648
EpiGII9344626Q	11	1017746	1017920	2.93E-07	rs10902233
EpiGII9386180Q	4	2486858	2487068	2.74E-06	rs6832191
EpiGII9388564Q	8	146055526	146055695	1.80E-06	8-146003327
EpiGII9402604AG	16	1504260	1504551	6.63E-07	rs35021217
EpiGII9448479Q	8	10797346	10797582	9.97E-07	8-10680107
EpiGII9658371Q	1	153631080	153631351	1.62E-06	rs1329101
EpiGII9659368Q	1	153939705	153940012	2.22E-06	rs7532808
EpiGII9680064Q	1	160011993	160012406	4.89E-08	rs16831234

.Chr: Chromosome. SNP rs ID: SNP rs ID or Chromosome-Position. Multiple rs IDs represent equally ranked SNPs

Appendix Table 2: CpGs associated with BMI p<10-7 in the primary study.

			Discovery			Replication		
CpG	**Chr-Position**	**Nearest gene**	**Effect (SE)**	**P**	**HetP**	**Effect (SE)**	**P**	**HetP**
cg06500161	21-43656587	*ABCG1*	34.8 (2.3)	6.0E-46	4.7E-2	19.9 (1.9)	1.1E-25	5.2E-3
cg00574958	11-68607622	*CPT1A*	-40.2 (3.1)	3.3E-33	8.2E-1	-21.3 (2.4)	2.8E-18	1.4E-1
cg12593793	1-156074135	*LMNA*	-32.0 (2.7)	3.0E-27	3.6E-1	-10.2 (1.7)	3.9E-9	1.4E-2
cg11024682	17-17730094	*SREBF1*	32.5 (3.0)	1.9E-23	7.9E-5	22.7 (2.2)	3.0E-25	6.0E-1
cg27243685	21-43642366	*ABCG1*	31.0 (3.1)	8.2E-21	3.4E-4	17.4 (2.5)	6.7E-12	8.1E-5
cg09349128	22-50327986	*CRELD2*	-24.4 (2.4)	1.7E-20	3.7E-3	-13.7 (1.9)	7.3E-13	1.2E-2
cg18181703	17-76354621	*SOCS3*	-18.0 (1.8)	2.8E-20	5.5E-1	-12.1 (1.5)	1.6E-16	4.7E-2
cg17501210	6-166970252	*RPS6KA2*	-18.4 (1.9)	8.2E-19	1.8E-3	-15.2 (1.6)	6.4E-21	3.1E-1
cg16594806	1-59473943	*PHBP3*	-24.9 (2.6)	2.3E-18	5.3E-5	-3.5 (1.4)	1.3E-2	5.2E-1
cg06164260	3-187454439	*BCL6*	-21.3 (2.3)	8.1E-18	9.5E-2	-13.0 (2.2)	4.9E-9	5.5E-2
cg14476101	1-120255992	*PHGDH*	-14.5 (1.6)	2.1E-17	5.2E-2	-8.9 (1.2)	9.4E-13	6.5E-1
cg02650017	17-47301614	*PHOSPHO1*	-28.6 (3.2)	1.9E-16	4.7E-1	-15.2 (2.7)	2.0E-8	2.2E-4
cg19750657	13-38935967	*UFM1*	23.9 (2.7)	2.2E-16	1.9E-6	10.8 (1.6)	4.7E-11	6.1E-1
cg09664445	17-2612406	*CLUH*	31.2 (3.5)	2.4E-16	5.4E-2	8.5 (2.2)	1.2E-4	5.7E-2
cg17260706	11-118782879	*BCL9L*	-28.4 (3.3)	1.2E-15	7.7E-2	-9.7 (2.1)	4.7E-6	6.0E-1
cg26403843	5-158634085	*RNF145*	10.7 (1.3)	2.9E-15	4.2E-1	7.7 (1.0)	5.1E-14	7.5E-1
cg11376147	11-57261198	*SLC43A1*	-25.3 (3.0)	5.8E-15	1.1E-4	-9.1 (2.4)	1.6E-4	7.2E-1
cg12484113	1-27898757	*AHDC1*	18.6 (2.2)	1.2E-14	1.8E-1	14.3 (2.4)	2.2E-9	4.6E-1
cg08857797	17-40927699	*VPS25*	19.8 (2.4)	3.3E-14	1.3E-5	3.9 (1.3)	2.3E-3	4.3E-2
cg15721584	3-181326755	*SOX2-OT*	10.4 (1.3)	4.9E-14	1.8E-4	6.4 (1.1)	1.6E-8	3.4E-1
cg08309687	21-35320596	*LINC00649*	-15.2 (1.9)	1.1E-13	4.4E-4	-4.6 (1.1)	2.9E-5	9.4E-1
cg06946797	16-11422409	*RMI2*	-14.5 (1.8)	1.5E-13	4.1E-2	-7.0 (1.4)	6.0E-7	8.6E-1

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg03523676	14-24540235	*CPNE6*	19.5 (2.5)	3.5E-13	1.7E-1	5.1 (1.6)	1.7E-3	6.2E-3
cg03725309	1-109757585	*SARS*	-24.8 (3.2)	3.8E-13	1.7E-1	-8.2 (1.6)	1.8E-7	5.3E-1
cg25217710	1-156609523	*BCAN*	26.9 (3.4)	4.4E-13	3.4E-1	5.8 (2.0)	3.4E-3	1.8E-1
cg16246545	1-120255941	*PHGDH*	-14.0 (1.8)	4.8E-13	3.3E-1	-7.5 (1.5)	3.9E-7	9.2E-1
cg16163382	2-37938640	*CDC42EP3*	-20.3 (2.6)	5.8E-13	7.8E-1	-6.5 (2.0)	1.2E-3	2.7E-1
cg18608055	19-1130866	*SBNO2*	-21.7 (2.8)	5.9E-13	1.0E-3	-5.8 (1.3)	4.7E-6	2.6E-5
cg10922280	16-68034227	*DUS2L*	23.8 (3.1)	6.1E-13	3.0E-2	7.7 (1.9)	4.2E-5	4.9E-1
cg00994936	19-1423902	*DAZAP1*	13.9 (1.8)	7.7E-13	4.8E-1	7.0 (1.6)	7.6E-6	1.3E-2
cg05778424	17-55169508	*AKAP1*	21.6 (2.8)	1.0E-12	3.6E-3	2.1 (1.2)	8.9E-2	7.6E-2
cg07769588	19-10655622	*ATG4D*	8.4 (1.1)	1.3E-12	1.8E-1	7.1 (1.7)	2.2E-5	9.8E-1
cg02716826	9-33447032	*AQP3*	-14.2 (1.9)	1.6E-12	7.3E-2	-10.2 (1.6)	3.5E-10	2.3E-1
cg01243823	16-50732212	*NOD2*	-13.8 (1.8)	1.7E-12	2.7E-2	-8.4 (1.2)	2.2E-12	7.5E-1
cg00244001	10-126336805	*FAM53B*	-33.3 (4.4)	2.0E-12	9.9E-1	-6.7 (1.9)	3.4E-4	4.9E-2
cg03318904	22-39801522	*TAB1*	22.3 (2.9)	2.2E-12	4.4E-3	8.3 (1.5)	2.6E-8	3.6E-2
cg23866916	19-1155738	*SBNO2*	-31.3 (4.1)	2.4E-12	3.8E-6	-5.1 (1.4)	2.5E-4	8.8E-2
cg24174557	17-57903544	*VMP1*	-11.6 (1.5)	2.4E-12	2.3E-1	-7.0 (1.3)	3.0E-8	1.3E-1
cg17058475	11-68607737	*CPT1A*	-18.2 (2.4)	2.8E-12	9.9E-3	-12.5 (2.0)	9.6E-10	2.3E-1
cg26253134	2-70751721	*TGFA*	-23.8 (3.2)	3.2E-12	2.7E-1	-7.4 (1.4)	2.3E-7	7.4E-5
cg22700686	1-153538764	*S100A2*	-25.9 (3.4)	3.2E-12	3.7E-2	-6.3 (1.7)	1.4E-4	1.1E-1
cg23232188	3-121556543	*EAF2*	18.7 (2.5)	4.0E-12	4.3E-2	5.4 (1.3)	1.9E-5	1.6E-2
cg25889918	2-179916184	*CCDC141*	17.0 (2.3)	6.0E-12	9.8E-1	0.4 (1.8)	8.5E-1	2.3E-1
cg18513344	3-195531298	*MUC4*	-14.7 (2.0)	6.0E-12	7.7E-1	-12.1 (2.3)	1.1E-7	5.5E-2
cg21486834	17-74477542	*RHBDF2*	17.9 (2.4)	6.4E-12	4.5E-1	5.3 (1.9)	6.2E-3	6.8E-3
cg22103219	7-101934892	*SH2B2*	-14.9 (2.0)	7.0E-12	9.5E-1	-5.0 (1.3)	2.0E-4	5.4E-1
cg16611584	17-19809078	*AKAP10*	7.1 (1.0)	1.0E-11	4.8E-1	3.3 (1.0)	6.3E-4	4.5E-1

			Discovery			Replication		
CpG	**Chr-Position**	**Nearest gene**	**Effect (SE)**	**P**	**HetP**	**Effect (SE)**	**P**	**HetP**
cg03327570	2-145304883	*ZEB2*	-18.5 (2.5)	1.0E-11	3.5E-1	-8.5 (1.5)	1.5E-8	7.5E-1
cg07573872	19-1126342	*SBNO2*	-15.7 (2.1)	1.3E-11	1.8E-8	-9.8 (1.3)	6.8E-14	5.6E-2
cg26663590	16-28959310	*NFATC2IP*	16.6 (2.3)	1.3E-11	5.2E-2	4.8 (1.2)	3.3E-5	6.9E-1
cg06603309	11-2724144	*KCNQ1*	-27.7 (3.8)	1.4E-11	2.5E-2	-4.2 (1.5)	5.3E-3	9.5E-1
cg11969813	17-79816559	*P4HB*	17.5 (2.4)	1.6E-11	2.0E-2	5.4 (1.3)	2.6E-5	1.4E-1
cg11202345	17-76976057	*LGALS3BP*	17.8 (2.5)	2.3E-11	7.8E-3	7.4 (1.6)	1.9E-6	1.4E-1
cg18098839	3-167742700	*GOLIM4*	-15.7 (2.2)	2.7E-11	2.1E-1	-14.1 (2.3)	1.7E-9	1.9E-2
cg02119938	15-78505051	*ACSBG1*	-23.4 (3.2)	2.9E-11	9.2E-1	-12.6 (2.1)	1.1E-9	7.2E-3
cg24678869	1-153919638	*DENND4B*	24.3 (3.4)	3.6E-11	5.9E-1	4.4 (2.4)	6.0E-2	5.0E-2
cg10927968	11-1807333	*CTSD*	12.8 (1.8)	3.9E-11	3.3E-1	6.6 (1.5)	6.0E-6	1.3E-1
cg07037944	15-64290807	*DAPK2*	-21.4 (3.0)	4.9E-11	3.3E-1	-9.8 (1.9)	1.3E-7	8.0E-1
cg02711608	19-47287964	*SLC1A5*	-17.2 (2.4)	5.6E-11	2.9E-1	-10.6 (1.8)	6.0E-9	5.9E-1
cg05648472	11-45232364	*PRDM11*	16.8 (2.4)	6.1E-11	3.8E-1	7.6 (2.1)	3.4E-4	1.2E-1
cg07136133	11-36422377	*PRR5L*	-19.8 (2.8)	6.5E-11	1.8E-2	-8.1 (1.6)	2.3E-7	1.7E-1
cg19420720	17-79816504	*P4HB*	21.6 (3.1)	7.3E-11	1.1E-1	4.4 (1.4)	1.6E-3	1.1E-1
cg24679890	19-17246356	*MYO9B*	22.3 (3.2)	8.3E-11	6.5E-1	4.8 (1.2)	1.0E-4	8.2E-2
cg00431050	10-103985730	*ELOVL3*	-26.8 (3.8)	1.1E-10	3.0E-3	-5.6 (1.9)	2.9E-3	3.4E-2
cg02286155	5-176826262	*SLC34A1*	23.2 (3.3)	1.2E-10	2.9E-1	9.4 (2.4)	9.3E-5	7.1E-1
cg14870271	17-76976010	*LGALS3BP*	16.4 (2.4)	1.3E-10	6.2E-2	9.8 (1.6)	2.0E-9	9.2E-2
cg09182678	22-50328711	*CRELD2*	-24.6 (3.5)	1.3E-10	2.1E-2	-9.3 (2.2)	1.9E-5	1.7E-2
cg26361535	8-144576604	*ZC3H3*	11.0 (1.6)	1.4E-10	3.8E-1	5.8 (1.3)	8.7E-6	5.3E-1
cg17901584	1-55353706	*DHCR24*	-13.7 (2.0)	1.4E-10	9.0E-7	-7.3 (1.2)	4.2E-10	1.0E-1
cg26033520	10-74004071	*ANAPC16*	11.9 (1.7)	1.6E-10	1.1E-2	8.5 (1.4)	2.9E-9	1.6E-1
cg11650298	13-44690989	*SMIM2-AS1*	-26.6 (3.8)	1.7E-10	2.8E-1	-10.6 (2.8)	1.3E-4	6.1E-2
cg10717869	1-205780912	*SLC41A1*	21.2 (3.1)	1.8E-10	4.9E-1	11.3 (2.1)	9.1E-8	4.3E-1

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg24843003	19-1409547	*DAZAP1*	13.9 (2.0)	2.0E-10	2.6E-1	7.2 (1.8)	9.8E-5	1.0E-1
cg04610187	17-76360794	*SOCS3*	-3.8 (0.6)	2.1E-10	1.1E-2	-4.9 (1.3)	1.5E-4	1.7E-3
cg19589396	8-103937374	*RPL5P24*	-14.3 (2.1)	2.3E-10	9.1E-1	-6.9 (1.3)	2.2E-7	5.4E-1
cg26804423	7-8201134	*ICA1*	20.5 (3.0)	2.4E-10	4.8E-1	4.7 (1.6)	2.6E-3	2.3E-3
cg09554443	1-167487762	*CD247*	-14.5 (2.1)	2.5E-10	9.0E-4	-4.9 (1.3)	2.5E-4	2.3E-1
cg22488164	12-14716910	*PLBD1*	14.4 (2.1)	2.5E-10	4.4E-1	6.6 (1.4)	5.2E-6	1.2E-2
cg27614723	15-92399897	*SLCO3A1*	24.5 (3.6)	2.7E-10	3.3E-2	3.9 (1.7)	2.1E-2	3.6E-3
cg12992827	3-101901234	*ZPLD1*	-13.0 (1.9)	2.9E-10	1.5E-4	-7.1 (1.5)	2.2E-6	2.0E-1
cg06690548	4-139162808	*SLC7A11*	-6.3 (0.9)	3.0E-10	7.6E-9	-11.7 (1.6)	1.0E-12	2.9E-1
cg01101459	1-234871477	*LINC00184*	14.2 (2.1)	3.1E-10	5.8E-1	5.7 (1.3)	1.3E-5	2.7E-2
cg24531955	8-23154691	*LOXL2*	-11.0 (1.6)	3.2E-10	8.1E-1	-9.4 (1.6)	9.1E-9	6.2E-1
cg07021906	16-87866833	*SLC7A5*	14.2 (2.1)	3.2E-10	3.2E-1	4.8 (1.4)	4.5E-4	5.6E-3
cg00851028	1-234905772	*RNY4P16*	23.3 (3.4)	3.3E-10	1.9E-1	11.0 (1.9)	7.9E-9	2.3E-1
cg22012981	3-58522689	*ACOX2*	10.2 (1.5)	3.5E-10	6.8E-2	8.1 (2.3)	5.8E-4	6.4E-1
cg00108715	3-52565015	*NT5DC2*	21.3 (3.2)	4.0E-10	1.2E-1	7.4 (2.0)	2.4E-4	7.3E-2
cg15323828	1-226053673	*TMEM63A*	-17.5 (2.6)	4.0E-10	8.4E-1	-5.0 (1.9)	8.5E-3	1.7E-3
cg06207201	16-50701535	*SNX20*	-18.0 (2.7)	4.6E-10	7.1E-2	-6.0 (2.1)	4.0E-3	6.8E-5
cg17178175	2-178109973	*NFE2L2*	-13.8 (2.0)	5.1E-10	6.2E-1	-6.0 (1.3)	2.9E-6	2.5E-1
cg25435714	7-157083381	*RN7SL142P*	16.5 (2.5)	5.3E-10	4.3E-2	3.4 (1.4)	1.5E-2	4.2E-1
cg19373099	2-210008092	*CRYGFP*	9.0 (1.3)	5.7E-10	1.3E-5	4.6 (1.1)	1.1E-5	1.4E-1
cg26357885	14-65006204	*HSPA2*	-21.8 (3.2)	5.7E-10	6.9E-1	-3.5 (1.5)	2.4E-2	5.2E-1
cg21429551	7-30635762	*GARS*	-6.9 (1.0)	6.2E-10	4.7E-1	-4.7 (0.9)	2.8E-7	6.4E-1
cg25096107	14-106037781	*IGHA2*	-23.2 (3.5)	7.0E-10	1.3E-1	-7.5 (2.5)	2.6E-3	3.9E-1
cg08648047	1-11028561	*C1orf127*	12.6 (1.9)	7.2E-10	2.9E-1	7.2 (2.2)	1.3E-3	2.7E-1
cg10919522	14-74227441	*ELMSAN1*	-14.1 (2.1)	7.5E-10	1.7E-1	-7.7 (1.6)	1.4E-6	7.4E-1

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg19695507	10-13526193	*BEND7*	9.8 (1.5)	7.8E-10	3.3E-1	4.2 (1.6)	9.3E-3	5.9E-1
cg10508317	17-76355146	*SOCS3*	-14.8 (2.2)	8.0E-10	8.3E-1	-9.9 (2.2)	5.0E-6	2.5E-2
cg11183227	15-91455407	*MAN2A2*	18.9 (2.8)	8.1E-10	6.8E-5	4.5 (1.4)	1.7E-3	8.0E-1
cg00441709	17-2528280	*PAFAH1B1*	12.4 (1.9)	8.1E-10	2.4E-1	4.0 (1.4)	2.9E-3	2.6E-1
cg10179300	5-14147618	*TRIO*	17.1 (2.6)	9.1E-10	2.7E-1	13.0 (2.6)	4.4E-7	3.8E-1
cg05095590	7-2139259	*MAD1L1*	9.5 (1.4)	9.6E-10	1.7E-1	5.1 (1.4)	4.1E-4	4.2E-1
cg05845030	12-91573247	*DCN*	-19.3 (2.9)	1.1E-9	3.1E-3	-5.4 (1.3)	5.9E-5	3.5E-1
cg03885055	1-16723232	*SPATA21*	-25.2 (3.8)	1.3E-9	8.6E-1	-6.4 (2.2)	3.8E-3	4.2E-1
cg11927233	5-170816542	*NPM1*	10.5 (1.6)	1.3E-9	6.8E-1	6.9 (1.5)	4.3E-6	5.6E-1
cg15138339	17-40715222	*COASY*	17.2 (2.6)	1.3E-9	1.2E-1	8.3 (1.9)	1.9E-5	7.4E-1
cg19998073	14-89078443	*ZC3H14*	18.6 (2.8)	1.3E-9	5.5E-1	5.8 (2.2)	9.2E-3	3.0E-1
cg10819350	19-10655686	*ATG4D*	13.6 (2.1)	1.3E-9	1.8E-2	5.1 (1.5)	6.0E-4	4.4E-1
cg27269962	7-127540997	*SND1*	22.3 (3.4)	1.5E-9	3.9E-1	6.8 (2.4)	4.0E-3	2.3E-1
cg07682160	19-18959935	*UPF1*	16.8 (2.6)	1.5E-9	1.1E-2	5.0 (1.5)	1.0E-3	3.0E-1
cg18120259	6-43894639	*C6orf223*	-12.6 (1.9)	1.6E-9	1.5E-3	-7.9 (1.3)	3.2E-9	2.9E-1
cg13997435	1-153538406	*S100A2*	-22.5 (3.4)	1.6E-9	2.0E-1	-11.1 (2.2)	3.3E-7	4.4E-1
cg03957124	6-37016869	*COX6A1P2*	-22.3 (3.4)	1.6E-9	1.5E-4	-3.1 (1.2)	8.5E-3	8.7E-2
cg07728579	15-83475013	*FSD2*	14.9 (2.3)	1.7E-9	9.5E-1	9.2 (1.9)	2.0E-6	1.4E-1
cg23068772	7-30737061	*CRHR2*	-14.7 (2.3)	2.0E-9	2.4E-1	-9.2 (1.8)	3.6E-7	3.4E-1
cg08726900	16-89550474	*ANKRD11*	-9.9 (1.5)	2.0E-9	1.3E-4	-3.0 (1.4)	3.0E-2	7.3E-1
cg27087650	19-45255796	*BCL3*	-21.5 (3.3)	2.1E-9	2.3E-1	-3.4 (1.6)	3.2E-2	4.7E-1
cg07202479	1-159174162	*DARC*	-14.5 (2.2)	2.2E-9	6.3E-1	-10.9 (2.0)	4.0E-8	2.4E-1
cg22441770	1-153929592	*CRTC2*	16.6 (2.6)	2.2E-9	6.7E-1	5.1 (1.7)	3.0E-3	4.2E-1
cg26894079	11-122954435	*CLMP*	-14.2 (2.2)	2.2E-9	2.9E-2	-8.9 (1.4)	3.8E-10	2.3E-1
cg19748495	1-150979359	*FAM63A*	-23.9 (3.7)	2.5E-9	3.4E-1	-2.4 (1.6)	1.2E-1	7.0E-1

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg22590032	5-180050565	*FLT4*	13.9 (2.1)	2.5E-9	5.3E-1	8.6 (2.2)	1.1E-4	1.2E-2
cg09831562	3-181327125	*SOX2-OT*	12.7 (2.0)	2.6E-9	4.8E-4	9.0 (1.8)	2.8E-7	3.7E-1
cg04577162	7-73667397	*RFC2*	20.3 (3.2)	2.6E-9	1.1E-1	6.0 (1.7)	2.7E-4	2.9E-1
cg08640498	12-57828515	*INHBC*	22.9 (3.6)	2.7E-9	8.1E-1	3.0 (2.7)	2.6E-1	1.6E-1
cg00220721	11-36422443	*PRR5L*	-21.6 (3.4)	2.7E-9	3.9E-2	-5.0 (1.7)	2.9E-3	2.7E-2
cg11080651	5-10445523	*ROPN1L*	-26.0 (4.1)	2.9E-9	8.6E-1	-4.5 (1.7)	9.1E-3	2.9E-3
cg03940776	6-158490013	*SYNJ2*	-25.9 (4.0)	3.0E-9	8.3E-2	-11.9 (2.1)	2.0E-8	5.8E-2
cg02203067	16-87866800	*SLC7A5*	17.9 (2.8)	3.0E-9	6.2E-1	3.1 (1.6)	5.0E-2	1.0E-2
cg00711896	16-30410051	*ZNF48*	20.5 (3.2)	3.1E-9	3.9E-1	6.8 (1.8)	2.0E-4	7.7E-1
cg08548559	22-31686097	*PIK3IP1*	-9.0 (1.4)	3.4E-9	7.5E-1	-6.9 (1.4)	4.9E-7	1.0E-1
cg13123009	6-31681882	*LY6G6F*	18.4 (2.9)	3.4E-9	4.7E-1	8.2 (1.9)	1.2E-5	5.0E-2
cg26952928	8-142230233	*SLC45A4*	18.1 (2.8)	3.6E-9	3.4E-1	4.8 (2.1)	2.1E-2	1.3E-1
cg01881899	21-43652704	*ABCG1*	23.1 (3.6)	3.6E-9	6.6E-1	10.4 (2.3)	3.9E-6	5.1E-3
cg13668129	19-41768075	*HNRNPUL1*	-26.7 (4.2)	3.7E-9	2.7E-1	-3.5 (2.6)	1.8E-1	7.6E-1
cg06012428	6-157477204	*ARID1B*	-20.8 (3.3)	3.7E-9	4.0E-1	-3.0 (1.3)	2.1E-2	4.2E-1
cg04011474	2-28904455	*RNA5SP89*	-17.3 (2.7)	3.7E-9	1.8E-1	-7.6 (1.6)	2.7E-6	2.9E-2
cg06560379	6-44231305	*NFKBIE*	-29.8 (4.7)	3.8E-9	2.6E-3	-16.6 (2.8)	2.5E-9	4.1E-3
cg06192883	15-52554171	*MYO5C*	15.8 (2.5)	3.9E-9	1.2E-7	9.3 (1.4)	8.7E-11	9.1E-1
cg17836612	17-76976357	*LGALS3BP*	16.0 (2.5)	4.3E-9	5.8E-1	8.5 (1.8)	1.7E-6	1.3E-1
cg27050612	17-46133198	*NFE2L1*	-33.2 (5.2)	4.3E-9	6.3E-2	-5.6 (1.6)	3.3E-4	1.4E-1
cg07504977	10-102131012	*LINC00263*	12.4 (2.0)	4.5E-9	1.6E-2	8.1 (1.3)	1.7E-10	6.1E-1
cg03433986	11-62477624	*BSCL2*	-18.6 (2.9)	4.5E-9	1.7E-1	-6.6 (1.8)	2.0E-4	7.9E-1
cg00417304	11-45124456	*PRDM11*	-2.6 (0.4)	4.9E-9	1.6E-2	-4.6 (1.5)	1.7E-3	1.2E-1
cg27037013	21-35320667	*LINC00649*	-9.0 (1.4)	5.3E-9	3.1E-2	-2.6 (1.0)	8.2E-3	9.1E-1
cg18219562	17-41773643	*MEOX1*	19.3 (3.1)	5.7E-9	6.9E-1	6.3 (1.8)	3.5E-4	1.3E-3

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg09581649	16-4572735	*CDIP1*	21.8 (3.5)	5.9E-9	2.7E-1	3.0 (2.3)	1.9E-1	3.7E-1
cg19566658	7-100466241	*TRIP6*	13.5 (2.2)	5.9E-9	7.2E-2	4.4 (1.5)	3.0E-3	2.1E-1
cg25570328	2-108903952	*SULT1C2*	-18.3 (2.9)	6.0E-9	2.6E-2	-3.7 (1.3)	5.6E-3	7.0E-1
cg01798813	17-3906674	*ZZEF1*	21.9 (3.5)	6.1E-9	7.5E-2	13.2 (3.0)	8.7E-6	8.1E-1
cg05063895	16-2073518	*SLC9A3R2*	-29.8 (4.7)	6.3E-9	4.9E-1	-7.1 (2.6)	5.8E-3	2.4E-1
cg08972190	7-2138995	*MAD1L1*	26.7 (4.3)	6.8E-9	2.7E-3	9.4 (1.7)	2.0E-8	5.8E-2
cg09613192	2-181388538	*FTH1P20*	10.6 (1.7)	7.3E-9	3.9E-1	6.2 (1.5)	2.9E-5	2.8E-1
cg02417857	3-181326706	*SOX2-OT*	9.6 (1.5)	7.6E-9	1.5E-2	4.8 (1.2)	5.3E-5	5.5E-1
ch.2.30415474F	2-30561970	*LBH*	-35.4 (5.7)	7.7E-9	5.0E-1	-12.4 (3.4)	2.3E-4	2.2E-1
cg27637521	17-76355202	*SOCS3*	-20.6 (3.3)	7.9E-9	1.9E-1	-10.5 (2.4)	1.8E-5	2.6E-2
cg10975897	6-15504844	*JARID2*	-12.9 (2.1)	8.2E-9	2.4E-1	-4.9 (1.3)	1.1E-4	8.5E-1
cg25649826	17-20938740	*USP22*	16.5 (2.7)	8.9E-9	5.3E-1	6.4 (1.8)	4.2E-4	5.9E-1
cg13274938	17-38493822	*RARA*	22.8 (3.7)	9.2E-9	3.1E-2	10.0 (2.2)	6.1E-6	1.8E-1
cg07471614	8-125855152	*LINC00964*	21.3 (3.4)	9.4E-9	1.2E-1	8.1 (2.8)	3.2E-3	3.9E-1
cg02079413	11-2986505	*NAP1L4*	14.6 (2.4)	9.4E-9	8.8E-1	7.4 (1.7)	2.1E-5	7.6E-1
cg06397161	22-39760059	*SYNGR1*	13.6 (2.2)	1.0E-8	7.3E-1	4.9 (1.4)	6.6E-4	9.8E-1
cg17320136	5-10567905	*ANKRD33B*	20.2 (3.3)	1.1E-8	2.3E-2	6.1 (2.0)	1.7E-3	1.6E-1
cg00547480	10-103875961	*LDB1*	22.0 (3.6)	1.1E-8	1.6E-1	4.2 (1.5)	4.4E-3	7.7E-3
cg00094412	6-29592854	*GABBR1*	-16.6 (2.7)	1.1E-8	4.3E-1	-8.2 (1.8)	3.0E-6	7.2E-1
cg27115863	22-37921640	*CARD10*	-13.6 (2.2)	1.1E-8	1.8E-3	-9.4 (1.5)	7.6E-10	4.0E-2
cg14264316	9-134280803	*PRRC2B*	12.5 (2.0)	1.1E-8	5.3E-1	5.0 (1.4)	3.9E-4	1.6E-2
cg07201717	1-25031947	*SRRM1*	-25.3 (4.1)	1.3E-8	3.3E-1	-1.3 (1.2)	2.8E-1	4.2E-1
cg11588197	11-128391494	*ETS1*	-11.9 (1.9)	1.3E-8	9.9E-3	-1.6 (1.3)	2.0E-1	3.5E-1
cg10734665	15-26107410	*ATP10A*	-13.2 (2.1)	1.3E-8	1.0E-2	-4.9 (1.6)	2.3E-3	4.0E-1
cg09777883	11-112093696	*BCO2*	15.7 (2.5)	1.3E-8	5.2E-1	4.3 (1.7)	1.1E-2	1.2E-2

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg25178683	17-76976267	*LGALS3BP*	15.0 (2.4)	1.3E-8	1.4E-2	8.3 (1.5)	3.4E-8	1.6E-1
cg11832534	1-3563998	*WRAP73*	12.4 (2.0)	1.3E-8	8.8E-1	10.3 (2.2)	1.6E-6	2.5E-1
cg26470501	19-45252955	*BCL3*	-15.6 (2.5)	1.3E-8	5.7E-2	-11.8 (1.9)	1.0E-9	5.9E-1
cg23032421	3-3152038	*IL5RA*	-18.8 (3.1)	1.4E-8	1.6E-1	-5.7 (1.5)	1.4E-4	6.6E-2
cg07012687	17-80195180	*SLC16A3*	14.0 (2.3)	1.4E-8	1.1E-1	6.5 (1.4)	2.4E-6	2.2E-3
cg13097800	14-47104140	*RPL10L*	-18.0 (2.9)	1.4E-8	7.0E-1	-4.4 (1.8)	1.3E-2	7.9E-1
cg05720226	7-116786597	*ST7*	18.9 (3.1)	1.4E-8	4.5E-2	8.0 (2.2)	2.1E-4	2.6E-1
cg10505902	1-144892111	*PDE4DIP*	-14.8 (2.4)	1.4E-8	3.4E-2	-10.7 (2.3)	3.2E-6	9.6E-2
cg02050917	1-2173571	*SKI*	14.5 (2.4)	1.5E-8	2.0E-1	1.8 (1.5)	2.3E-1	5.4E-1
cg09559780	16-50727221	*NOD2*	-20.2 (3.3)	1.5E-8	7.9E-1	-4.8 (1.8)	8.6E-3	7.9E-1
cg26836479	19-42706353	*DEDD2*	-25.7 (4.2)	1.5E-8	8.5E-1	-5.1 (1.9)	6.5E-3	3.5E-1
cg10513161	3-183705727	*ABCC5*	22.4 (3.7)	1.5E-8	3.6E-1	7.5 (2.3)	1.2E-3	1.1E-2
cg00144180	2-240294362	*HDAC4*	15.4 (2.5)	1.5E-8	1.2E-1	6.8 (1.2)	1.5E-8	4.1E-1
cg25616160	17-39655863	*KRT13*	27.7 (4.5)	1.6E-8	3.7E-1	5.1 (3.1)	9.4E-2	5.2E-1
cg15357118	2-128927972	*UGGT1*	15.3 (2.5)	1.6E-8	3.1E-1	10.8 (1.9)	1.3E-8	5.9E-2
cg22534374	1-201511610	*RPS10P7*	-7.3 (1.2)	1.7E-8	5.2E-1	-5.1 (1.2)	2.4E-5	2.2E-1
cg01176028	21-43653234	*ABCG1*	9.0 (1.5)	1.7E-8	1.3E-2	5.4 (1.5)	2.1E-4	1.6E-1
cg21037180	8-82276987	*PMP2*	2.8 (0.5)	1.8E-8	2.5E-1	1.1 (1.1)	3.2E-1	2.4E-1
cg06898549	12-41083590	*CNTN1*	9.8 (1.6)	1.8E-8	1.7E-1	5.7 (1.3)	7.0E-6	1.0E-1
cg21766592	19-47288066	*SLC1A5*	-16.6 (2.7)	1.8E-8	8.4E-1	-6.4 (1.4)	9.5E-6	2.4E-1
cg23261443	7-101596340	*CUX1*	-15.5 (2.5)	1.9E-8	5.1E-1	-3.7 (1.3)	4.3E-3	2.6E-2
cg19618279	17-40715228	*COASY*	15.7 (2.6)	1.9E-8	7.6E-1	6.7 (1.9)	4.7E-4	8.0E-1
cg12272104	19-1423967	*DAZAP1*	19.5 (3.2)	2.0E-8	4.8E-1	2.5 (2.3)	2.6E-1	5.4E-2
cg00634542	2-219254588	*SLC11A1*	15.6 (2.6)	2.1E-8	5.4E-1	4.8 (1.7)	5.9E-3	7.7E-2
cg10636246	1-159046973	*AIM2*	-12.7 (2.1)	2.1E-8	1.6E-2	-6.5 (1.3)	3.4E-7	4.6E-1

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg24469729	7-27160520	*HOXA-AS2*	16.2 (2.7)	2.3E-8	5.6E-2	7.4 (1.9)	8.6E-5	1.2E-1
cg00973118	16-374570	*AXIN1*	9.8 (1.6)	2.3E-8	9.2E-1	6.7 (1.5)	1.6E-5	3.1E-1
cg00863378	16-56549757	*BBS2*	14.8 (2.4)	2.4E-8	4.1E-1	4.9 (1.8)	5.8E-3	6.7E-2
cg03078551	17-41656298	*ETV4*	-20.6 (3.4)	2.4E-8	7.9E-4	-7.5 (2.1)	3.7E-4	3.8E-1
cg06559575	12-53490352	*IGFBP6*	-21.1 (3.5)	2.5E-8	1.9E-4	-7.4 (1.4)	2.0E-7	4.3E-1
cg26687842	13-41055491	*LINC00598*	14.2 (2.4)	2.6E-8	7.9E-1	7.4 (1.8)	3.5E-5	5.0E-1
cg12332902	2-98330020	*ZAP70*	-21.8 (3.6)	2.7E-8	2.2E-2	-1.3 (1.9)	4.8E-1	9.6E-1
cg22898082	19-11074428	*SMARCA4*	19.0 (3.2)	2.7E-8	2.4E-1	4.1 (1.5)	6.2E-3	2.1E-4
cg03050965	1-101705237	*S1PR1*	-29.5 (4.9)	3.0E-8	9.1E-2	-3.6 (1.6)	2.4E-2	7.2E-2
cg04232128	5-138861241	*TMEM173*	-23.0 (3.8)	3.0E-8	1.0E-1	-6.5 (1.6)	5.0E-5	8.1E-2
cg04126866	10-85932763	*C10orf99*	27.4 (4.6)	3.0E-8	4.7E-1	7.2 (2.2)	1.3E-3	8.5E-1
cg09697999	7-733198	*PRKAR1B*	10.0 (1.7)	3.1E-8	1.4E-1	1.6 (1.1)	1.3E-1	8.7E-1
cg25197194	3-128758787	*EFCC1*	-11.0 (1.8)	3.1E-8	3.5E-2	-3.6 (1.3)	4.4E-3	2.4E-1
cg10937494	17-80195101	*SLC16A3*	19.2 (3.2)	3.2E-8	8.7E-1	5.9 (1.9)	1.4E-3	2.0E-2
cg09315878	1-1152580	*SDF4*	-10.3 (1.7)	3.2E-8	2.4E-1	-4.5 (1.3)	3.9E-4	1.5E-1
cg16466652	19-6271960	*MLLT1*	17.4 (2.9)	3.3E-8	3.6E-1	3.5 (2.1)	1.0E-1	6.6E-2
cg08305942	16-79692354	*MAF*	-15.3 (2.6)	3.3E-8	7.8E-1	-13.0 (1.9)	1.3E-11	7.3E-1
cg13467672	17-40464625	*STAT5A*	-20.3 (3.4)	3.4E-8	7.4E-1	-0.7 (2.6)	7.8E-1	2.5E-1
cg20820107	1-12193099	*TNFRSF8*	-22.5 (3.8)	3.5E-8	6.4E-2	-3.6 (2.1)	8.4E-2	7.2E-3
cg03159676	16-85600536	*GSE1*	2.1 (0.4)	3.5E-8	3.5E-2	7.3 (1.4)	4.5E-7	6.5E-1
cg19881557	14-20967426	*RNASE10*	9.8 (1.6)	3.7E-8	2.3E-2	7.2 (1.5)	9.4E-7	3.5E-1
cg25230327	8-82042993	*PAG1*	-12.0 (2.0)	3.7E-8	6.0E-1	-3.5 (1.1)	1.6E-3	1.4E-1
cg18217136	20-36157651	*PPIAP3*	19.8 (3.3)	3.8E-8	6.4E-1	7.9 (2.4)	1.1E-3	4.2E-1
cg06180099	1-150979364	*FAM63A*	-17.8 (3.0)	4.0E-8	7.4E-1	-2.2 (1.5)	1.4E-1	6.2E-1
cg04524040	19-4153364	*CREB3L3*	-13.5 (2.3)	4.1E-8	3.8E-1	-3.7 (1.4)	7.4E-3	4.4E-1

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg08443038	16-89006877	*CBFA2T3*	-14.3 (2.4)	4.2E-8	5.3E-1	-5.0 (2.3)	3.2E-2	3.2E-1
cg13781414	9-138951648	*NACC2*	-12.0 (2.0)	4.3E-8	9.7E-1	-8.3 (2.4)	5.5E-4	9.9E-1
cg26878209	10-112375475	*SMC3*	14.9 (2.5)	4.3E-8	6.8E-4	5.2 (1.5)	8.0E-4	1.0x10+00
cg19217955	17-7123994	*ACADVL*	-14.8 (2.5)	4.3E-8	9.0E-2	-14.4 (3.9)	2.4E-4	4.7E-2
cg00871207	14-89080250	*EML5*	10.2 (1.7)	4.4E-8	4.9E-1	10.7 (2.4)	6.3E-6	6.5E-1
cg20600850	12-6712515	*CHD4*	-23.4 (4.0)	4.5E-8	4.3E-1	-4.3 (2.8)	1.2E-1	4.0E-2
cg04407417	12-52228651	*SCN8A*	21.1 (3.6)	4.5E-8	4.8E-3	1.5 (2.8)	6.0E-1	4.8E-1
cg00138407	3-47386505	*KLHL18*	14.1 (2.4)	4.5E-8	2.2E-1	3.6 (1.4)	8.4E-3	6.1E-1
cg27117792	12-102330180	*DRAM1*	-12.9 (2.2)	4.5E-8	2.8E-1	-5.3 (1.2)	5.1E-6	3.4E-1
cg13276570	5-10567643	*ANKRD33B*	7.2 (1.2)	4.5E-8	1.4E-1	9.4 (2.4)	1.2E-4	9.5E-1
cg14020176	17-72764985	*SLC9A3R1*	17.5 (3.0)	4.6E-8	3.7E-1	5.4 (1.8)	3.7E-3	9.4E-1
cg17971578	1-36852463	*STK40*	-12.9 (2.2)	4.6E-8	5.6E-4	-5.8 (1.7)	4.4E-4	4.1E-1
cg02560388	2-11969958	*LPIN1*	-11.4 (1.9)	4.7E-8	9.4E-1	-6.4 (1.7)	1.3E-4	4.7E-3
cg09152259	2-128156114	*MAP3K2*	-10.3 (1.7)	4.7E-8	3.5E-4	-11.2 (1.4)	4.1E-15	3.9E-2
cg13511324	14-104056883	*APOPT1*	11.6 (2.0)	4.8E-8	2.4E-1	1.0 (1.3)	4.6E-1	8.1E-1
cg13922488	19-14545201	*PKN1*	21.2 (3.6)	4.8E-8	2.5E-1	13.2 (2.8)	2.5E-6	4.9E-1
cg10970124	6-31634602	*CSNK2B*	19.5 (3.3)	4.9E-8	3.0E-1	9.7 (2.6)	2.0E-4	7.9E-1
cg01304182	16-30409908	*ZNF48*	16.1 (2.7)	5.0E-8	3.6E-1	3.2 (1.7)	6.1E-2	8.4E-1
cg09222732	6-466893	*EXOC2*	-26.2 (4.4)	5.0E-8	3.9E-1	-11.4 (2.7)	2.1E-5	5.1E-1
cg21108085	11-44591098	*CD82*	-18.1 (3.1)	5.0E-8	9.3E-1	-4.8 (1.3)	2.3E-4	3.0E-1
cg10814005	14-91711041	*GPR68*	-14.1 (2.4)	5.1E-8	5.9E-1	-5.7 (2.1)	5.4E-3	2.3E-2
cg10192877	21-43641690	*ABCG1*	23.7 (4.0)	5.1E-8	6.0E-4	8.1 (3.3)	1.2E-2	8.1E-1
cg26610247	8-142297175	*SLC45A4*	12.1 (2.1)	5.1E-8	1.8E-4	6.3 (1.5)	1.8E-5	8.8E-1
cg00673344	3-156807691	*LINC00880*	-16.0 (2.7)	5.3E-8	3.4E-1	-4.4 (1.4)	1.5E-3	8.9E-2
cg15681239	3-38080203	*DLEC1*	-11.3 (1.9)	5.5E-8	2.8E-1	-5.1 (1.4)	3.3E-4	1.2E-1

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg14559176	9-138872899	*UBAC1*	7.0 (1.2)	5.5E-8	3.5E-1	5.2 (2.6)	4.5E-2	1.1E-1
cg09276842	1-1155026	*SDF4*	-29.6 (5.0)	5.6E-8	5.1E-2	-6.5 (2.4)	6.4E-3	9.5E-1
cg22143698	5-10608058	*ANKRD33B*	15.2 (2.6)	5.7E-8	9.2E-1	2.7 (1.1)	1.2E-2	3.6E-2
cg24403644	20-42574624	*TOX2*	26.5 (4.5)	5.9E-8	3.5E-1	14.5 (2.3)	2.7E-10	5.3E-2
cg16815882	1-35908609	*KIAA0319L*	21.2 (3.6)	6.0E-8	1.7E-3	3.6 (1.5)	1.8E-2	1.5E-1
cg27547344	1-43765617	*TIE1*	18.3 (3.1)	6.2E-8	5.0E-2	5.9 (2.0)	3.7E-3	2.7E-2
cg22695339	17-7791630	*CHD3*	-16.1 (2.8)	6.3E-8	6.9E-1	-10.6 (2.9)	2.0E-4	6.0E-1
cg03747456	12-52586127	*KRT80*	-19.4 (3.3)	6.3E-8	9.6E-1	-7.6 (1.9)	5.1E-5	7.1E-2
cg25392060	8-142297121	*SLC45A4*	15.1 (2.6)	6.4E-8	3.5E-6	4.8 (1.6)	3.0E-3	2.8E-1
cg10438589	4-14531493	*LINC00504*	13.2 (2.3)	6.5E-8	3.8E-2	4.1 (1.1)	2.4E-4	1.2E-2
cg16395997	1-3562798	*WRAP73*	15.1 (2.6)	6.6E-8	3.0E-1	3.7 (1.4)	6.6E-3	5.8E-1
cg01419914	17-79374691	*MIR4740*	11.3 (1.9)	6.7E-8	8.6E-2	5.1 (1.4)	3.5E-4	2.6E-5
cg16846518	3-128062608	*EEFSEC*	-15.4 (2.6)	6.7E-8	4.9E-1	-4.8 (1.5)	1.8E-3	4.8E-2
cg27184903	15-29285727	*APBA2*	15.9 (2.7)	6.8E-8	1.9E-1	7.7 (2.3)	8.0E-4	1.6E-1
cg14902389	2-99342685	*MGAT4A*	-15.9 (2.7)	7.5E-8	7.2E-2	-5.2 (2.9)	6.9E-2	9.7E-1
cg24776142	6-16513829	*ATXN1*	14.6 (2.5)	7.6E-8	2.6E-1	1.9 (1.1)	9.4E-2	2.0E-1
cg08813944	17-71258589	*CPSF4L*	14.0 (2.4)	7.6E-8	6.0E-2	5.4 (1.8)	2.4E-3	5.4E-1
cg10549088	3-64277154	*PRICKLE2*	10.5 (1.8)	7.7E-8	3.3E-1	4.1 (2.0)	4.1E-2	9.9E-2
cg06988336	16-89922397	*SPIRE2*	15.3 (2.6)	7.8E-8	4.9E-1	3.9 (1.8)	3.2E-2	9.7E-3
cg01511901	13-31004719	*UBE2L5P*	-17.5 (3.0)	8.0E-8	2.7E-3	-5.2 (1.5)	3.3E-4	4.4E-2
cg25001190	1-61668835	*NFIA*	-10.8 (1.9)	8.1E-8	2.4E-1	-6.2 (1.5)	4.6E-5	3.3E-2
cg22650271	22-39760165	*SYNGR1*	20.5 (3.5)	8.2E-8	1.5E-1	5.1 (2.2)	1.9E-2	3.4E-1
cg26542660	4-56813860	*CEP135*	-22.1 (3.8)	8.5E-8	6.2E-1	-5.2 (1.8)	4.4E-3	4.8E-2
cg16578636	10-92987457	*PCGF5*	-7.0 (1.2)	8.8E-8	9.9E-3	-2.5 (1.2)	3.9E-2	1.2E-2
cg11614585	20-897050	*ANGPT4*	16.1 (2.8)	9.0E-8	1.6E-1	5.8 (2.3)	1.1E-2	4.1E-1

CpG	Chr-Position	Nearest gene	Discovery Effect (SE)	P	HetP	Replication Effect (SE)	P	HetP
cg13591783	9-75768868	*ANXA1*	-14.3 (2.5)	9.0E-8	7.4E-1	-4.5 (1.1)	2.3E-5	1.5E-1
cg00238353	10-129785537	*PTPRE*	10.2 (1.8)	9.2E-8	4.4E-1	5.1 (1.7)	2.4E-3	2.5E-1
cg17054691	17-79813439	*P4HB*	17.3 (3.0)	9.4E-8	7.5E-1	8.3 (2.3)	2.2E-4	2.4E-2
cg16778405	16-653443	*RAB40C*	29.3 (5.1)	9.8E-8	7.8E-1	28.3 (4.2)	1.9E-11	6.3E-1
cg04927537	17-76976091	*LGALS3BP*	10.6 (1.8)	9.9E-8	6.6E-4	7.6 (1.3)	9.9E-10	5.0E-2
cg08881019	1-156096178	*LMNA*	-24.0 (4.2)	1.0E-7	5.3E-1	-9.1 (2.4)	1.6E-4	9.6E-3

Chr: Chromosome. Discovery Effect (SE): Effect (SE) of the CpG effect in the meta-analyses, P-value: p-value of the CpG effect in the meta-analyses, HetP: Heterogeneity p-value (Chochran's q-test) of the CpG effect in the meta-analyses. Red: non-replicating CpGs

Appendix Table 3: Association of 187 CpG sites with clinical phenotypes

CpG Site	Trait	Effect (SE)	p-value	Trait	Effect (SE)	p-value
cg00094412	Chol	-0.010(0.0045)	2.76E-02	Ins	0.088(0.0043)	3.18E-85
cg00108715	Chol	-0.010(0.0045)	3.13E-02	Ins	0.087(0.0043)	1.24E-84
cg00138407	Chol	-0.009(0.0045)	4.03E-02	Ins	0.087(0.0043)	1.55E-84
cg00144180	Chol	-0.010(0.0045)	2.58E-02	Ins	0.087(0.0043)	9.50E-85
cg00238353	Chol	-0.011(0.0045)	1.91E-02	Ins	0.088(0.0043)	2.13E-85
cg00244001	Chol	-0.010(0.0045)	2.15E-02	Ins	0.088(0.0043)	5.61E-85
cg00431050	Chol	-0.011(0.0045)	1.58E-02	Ins	0.088(0.0043)	1.76E-84
cg00574958	Chol	-0.015(0.0045)	6.57E-04	Ins	0.086(0.0044)	2.23E-81
cg00634542	Chol	-0.010(0.0045)	2.47E-02	Ins	0.088(0.0043)	5.21E-85
cg00673344	Chol	-0.010(0.0045)	2.40E-02	Ins	0.088(0.0043)	1.67E-86
cg00711896	Chol	-0.011(0.0045)	1.80E-02	Ins	0.088(0.0043)	1.45E-85
cg00863378	Chol	-0.009(0.0045)	3.93E-02	Ins	0.087(0.0043)	1.75E-84
cg00973118	Chol	-0.010(0.0045)	2.17E-02	Ins	0.087(0.0043)	9.55E-84
cg01101459	Chol	-0.009(0.0045)	3.85E-02	Ins	0.087(0.0043)	1.88E-83
cg01243823	Chol	-0.011(0.0045)	1.48E-02	Ins	0.088(0.0043)	1.17E-85
cg01511901	Chol	-0.009(0.0045)	4.23E-02	Ins	0.088(0.0043)	7.96E-85
cg01798813	Chol	-0.008(0.0045)	7.59E-02	Ins	0.087(0.0043)	5.41E-83
cg02119938	Chol	-0.010(0.0045)	2.55E-02	Ins	0.087(0.0043)	1.70E-84
cg02286155	Chol	-0.010(0.0045)	2.55E-02	Ins	0.088(0.0043)	2.10E-84
cg02560388	Chol	-0.009(0.0045)	4.21E-02	Ins	0.087(0.0043)	1.20E-83
cg02650017	Chol	-0.007(0.0045)	1.03E-01	Ins	0.086(0.0043)	6.56E-81
cg02711608	Chol	-0.010(0.0045)	2.52E-02	Ins	0.088(0.0043)	5.67E-86
cg02716826	Chol	-0.009(0.0045)	5.39E-02	Ins	0.086(0.0043)	5.86E-82
cg03050965	Chol	-0.010(0.0045)	2.18E-02	Ins	0.088(0.0043)	1.02E-84
cg03159676	Chol	-0.009(0.0045)	4.12E-02	Ins	0.086(0.0043)	3.12E-82
cg03318904	Chol	-0.011(0.0045)	1.46E-02	Ins	0.086(0.0043)	2.03E-82
cg03327570	Chol	-0.009(0.0045)	3.77E-02	Ins	0.087(0.0043)	5.28E-84
cg03433986	Chol	-0.010(0.0045)	2.08E-02	Ins	0.088(0.0043)	2.34E-85
cg03523676	Chol	-0.009(0.0045)	5.92E-02	Ins	0.087(0.0043)	3.33E-84
cg03725309	Chol	-0.011(0.0045)	1.89E-02	Ins	0.087(0.0043)	6.46E-85
cg03885055	Chol	-0.010(0.0045)	3.38E-02	Ins	0.088(0.0043)	5.85E-85
cg03940776	Chol	-0.009(0.0045)	3.73E-02	Ins	0.087(0.0043)	3.08E-84
cg03957124	Chol	-0.010(0.0045)	3.09E-02	Ins	0.086(0.0043)	1.62E-81
cg04011474	Chol	-0.010(0.0045)	2.37E-02	Ins	0.087(0.0043)	1.05E-84
cg04126866	Chol	-0.010(0.0045)	2.18E-02	Ins	0.087(0.0043)	3.98E-84
cg04232128	Chol	-0.010(0.0045)	2.53E-02	Ins	0.087(0.0043)	1.93E-84
cg04524040	Chol	-0.010(0.0045)	2.15E-02	Ins	0.088(0.0043)	7.47E-86
cg04577162	Chol	-0.010(0.0045)	2.98E-02	Ins	0.087(0.0043)	1.43E-84
cg05063895	Chol	-0.011(0.0045)	1.54E-02	Ins	0.088(0.0043)	3.54E-86
cg05095590	Chol	-0.010(0.0045)	2.36E-02	Ins	0.088(0.0043)	1.27E-85
cg05648472	Chol	-0.009(0.0045)	5.57E-02	Ins	0.087(0.0043)	1.03E-82
cg05720226	Chol	-0.010(0.0045)	3.18E-02	Ins	0.087(0.0043)	2.50E-84
cg05845030	Chol	-0.010(0.0045)	2.75E-02	Ins	0.088(0.0043)	1.67E-84
cg06012428	Chol	-0.010(0.0045)	2.63E-02	Ins	0.088(0.0043)	2.33E-85
cg06164260	Chol	-0.010(0.0045)	2.05E-02	Ins	0.086(0.0043)	2.55E-82

cg06192883	Chol	-0.010(0.0045)	3.35E-02	Ins	0.085(0.0043)	3.30E-80
cg06500161	Chol	-0.005(0.0045)	2.29E-01	Ins	0.084(0.0044)	8.31E-78
cg06559575	Chol	-0.010(0.0045)	2.56E-02	Ins	0.087(0.0043)	4.21E-84
cg06603309	Chol	-0.009(0.0045)	5.26E-02	Ins	0.086(0.0043)	2.02E-82
cg06690548	Chol	-0.011(0.0045)	1.81E-02	Ins	0.088(0.0043)	4.76E-86
cg06898549	Chol	-0.011(0.0045)	1.89E-02	Ins	0.086(0.0043)	2.29E-83
cg06946797	Chol	-0.010(0.0045)	3.45E-02	Ins	0.086(0.0043)	5.41E-82
cg07021906	Chol	-0.011(0.0045)	1.59E-02	Ins	0.088(0.0043)	3.59E-85
cg07037944	Chol	-0.010(0.0045)	2.56E-02	Ins	0.087(0.0043)	6.16E-84
cg07136133	Chol	-0.010(0.0045)	2.30E-02	Ins	0.088(0.0043)	7.78E-85
cg07202479	Chol	-0.011(0.0045)	1.62E-02	Ins	0.087(0.0043)	1.76E-84
cg07471614	Chol	-0.011(0.0045)	1.75E-02	Ins	0.087(0.0043)	9.35E-84
cg07504977	Chol	-0.010(0.0045)	2.20E-02	Ins	0.087(0.0043)	4.06E-84
cg07682160	Chol	-0.009(0.0045)	4.58E-02	Ins	0.086(0.0043)	9.43E-82
cg07728579	Chol	-0.010(0.0045)	2.77E-02	Ins	0.088(0.0043)	2.65E-85
cg07769588	Chol	-0.008(0.0045)	7.04E-02	Ins	0.087(0.0043)	5.91E-83
cg08305942	Chol	-0.009(0.0045)	3.53E-02	Ins	0.087(0.0043)	6.88E-85
cg08309687	Chol	-0.011(0.0045)	1.94E-02	Ins	0.087(0.0043)	1.62E-83
cg08443038	Chol	-0.010(0.0045)	2.38E-02	Ins	0.087(0.0043)	1.17E-84
cg08548559	Chol	-0.011(0.0045)	1.56E-02	Ins	0.088(0.0043)	1.02E-85
cg08648047	Chol	-0.010(0.0045)	2.87E-02	Ins	0.087(0.0043)	1.19E-83
cg08726900	Chol	-0.011(0.0045)	1.36E-02	Ins	0.088(0.0043)	8.21E-85
cg08813944	Chol	-0.010(0.0045)	2.03E-02	Ins	0.087(0.0043)	1.56E-84
cg08857797	Chol	-0.011(0.0045)	1.22E-02	Ins	0.087(0.0043)	5.22E-84
cg09152259	Chol	-0.009(0.0045)	3.98E-02	Ins	0.087(0.0043)	1.40E-84
cg09222732	Chol	-0.010(0.0045)	3.11E-02	Ins	0.087(0.0043)	3.69E-84
cg09315878	Chol	-0.010(0.0045)	2.18E-02	Ins	0.088(0.0043)	2.49E-86
cg09349128	Chol	-0.006(0.0045)	1.57E-01	Ins	0.085(0.0044)	2.05E-79
cg09554443	Chol	-0.010(0.0045)	2.04E-02	Ins	0.087(0.0043)	7.70E-84
cg09613192	Chol	-0.010(0.0045)	2.42E-02	Ins	0.088(0.0043)	6.67E-85
cg09664445	Chol	-0.010(0.0045)	3.20E-02	Ins	0.087(0.0043)	3.93E-83
cg09777883	Chol	-0.010(0.0045)	1.95E-02	Ins	0.087(0.0043)	1.41E-84
cg10179300	Chol	-0.010(0.0045)	3.11E-02	Ins	0.087(0.0043)	7.87E-83
cg10438589	Chol	-0.010(0.0045)	3.28E-02	Ins	0.087(0.0043)	5.79E-84
cg10505902	Chol	-0.009(0.0045)	3.96E-02	Ins	0.087(0.0043)	6.90E-85
cg10513161	Chol	-0.009(0.0045)	3.63E-02	Ins	0.087(0.0043)	1.40E-84
cg10549088	Chol	-0.009(0.0045)	4.05E-02	Ins	0.087(0.0043)	6.41E-83
cg10717869	Chol	-0.010(0.0045)	2.19E-02	Ins	0.087(0.0043)	2.22E-84
cg10734665	Chol	-0.008(0.0045)	6.25E-02	Ins	0.087(0.0043)	5.15E-84
cg10814005	Chol	-0.010(0.0045)	2.57E-02	Ins	0.087(0.0043)	4.48E-84
cg10919522	Chol	-0.009(0.0045)	3.83E-02	Ins	0.087(0.0043)	5.31E-84
cg10922280	Chol	-0.010(0.0045)	3.27E-02	Ins	0.087(0.0043)	3.73E-83
cg10927968	Chol	-0.011(0.0045)	1.08E-02	Ins	0.087(0.0043)	4.17E-83
cg10975897	Chol	-0.011(0.0045)	1.84E-02	Ins	0.088(0.0043)	6.69E-86
cg11024682	Chol	-0.011(0.0045)	1.84E-02	Ins	0.086(0.0043)	2.11E-80
cg11080651	Chol	-0.010(0.0045)	2.11E-02	Ins	0.087(0.0043)	3.72E-84
cg11183227	Chol	-0.010(0.0045)	2.48E-02	Ins	0.087(0.0043)	5.01E-84
cg11202345	Chol	-0.010(0.0045)	3.05E-02	Ins	0.088(0.0043)	1.78E-84

cg11376147	Chol	-0.011(0.0045)	1.41E-02	Ins	0.088(0.0043)	2.35E-85
cg11614585	Chol	-0.010(0.0045)	2.11E-02	Ins	0.087(0.0043)	1.29E-83
cg11650298	Chol	-0.009(0.0045)	4.08E-02	Ins	0.087(0.0043)	3.73E-84
cg11832534	Chol	-0.009(0.0045)	3.56E-02	Ins	0.087(0.0043)	2.56E-83
cg11927233	Chol	-0.010(0.0045)	2.20E-02	Ins	0.088(0.0043)	2.62E-86
cg11969813	Chol	-0.010(0.0045)	2.83E-02	Ins	0.087(0.0043)	9.30E-84
cg12484113	Chol	-0.011(0.0045)	1.77E-02	Ins	0.088(0.0043)	3.34E-84
cg12593793	Chol	-0.011(0.0045)	1.26E-02	Ins	0.086(0.0043)	2.40E-81
cg12992827	Chol	-0.008(0.0045)	7.07E-02	Ins	0.087(0.0043)	5.87E-84
cg13097800	Chol	-0.010(0.0045)	2.51E-02	Ins	0.088(0.0043)	1.44E-86
cg13123009	Chol	-0.011(0.0045)	1.58E-02	Ins	0.087(0.0043)	5.97E-85
cg13274938	Chol	-0.010(0.0045)	2.94E-02	Ins	0.087(0.0043)	3.10E-83
cg13591783	Chol	-0.010(0.0045)	2.52E-02	Ins	0.088(0.0043)	4.36E-86
cg13781414	Chol	-0.009(0.0045)	4.63E-02	Ins	0.087(0.0043)	2.98E-83
cg13922488	Chol	-0.009(0.0045)	3.62E-02	Ins	0.087(0.0043)	1.58E-84
cg14020176	Chol	-0.010(0.0045)	2.03E-02	Ins	0.088(0.0043)	3.15E-85
cg14264316	Chol	-0.010(0.0045)	3.40E-02	Ins	0.088(0.0043)	3.00E-86
cg14476101	Chol	-0.011(0.0045)	1.73E-02	Ins	0.088(0.0043)	8.77E-85
cg15323828	Chol	-0.010(0.0045)	2.59E-02	Ins	0.087(0.0043)	7.30E-85
cg15357118	Chol	-0.009(0.0045)	3.72E-02	Ins	0.087(0.0043)	7.66E-84
cg15681239	Chol	-0.010(0.0045)	3.29E-02	Ins	0.087(0.0043)	1.10E-84
cg15721584	Chol	-0.009(0.0045)	4.15E-02	Ins	0.087(0.0043)	3.28E-82
cg16163382	Chol	-0.009(0.0045)	3.61E-02	Ins	0.086(0.0043)	1.03E-81
cg16578636	Chol	-0.009(0.0045)	3.94E-02	Ins	0.088(0.0043)	1.38E-85
cg16594806	Chol	-0.009(0.0045)	4.40E-02	Ins	0.087(0.0044)	3.42E-82
cg16611584	Chol	-0.010(0.0045)	2.00E-02	Ins	0.086(0.0043)	1.56E-82
cg16815882	Chol	-0.010(0.0045)	3.33E-02	Ins	0.087(0.0043)	1.04E-83
cg16846518	Chol	-0.010(0.0045)	2.87E-02	Ins	0.088(0.0043)	1.51E-85
cg17178175	Chol	-0.009(0.0045)	4.83E-02	Ins	0.087(0.0043)	1.43E-84
cg17260706	Chol	-0.011(0.0045)	1.71E-02	Ins	0.087(0.0043)	6.55E-83
cg17501210	Chol	-0.008(0.0045)	7.87E-02	Ins	0.086(0.0043)	2.69E-81
cg17901584	Chol	-0.006(0.0045)	2.08E-01	Ins	0.087(0.0043)	2.15E-83
cg17971578	Chol	-0.010(0.0045)	2.61E-02	Ins	0.088(0.0043)	5.84E-86
cg18098839	Chol	-0.009(0.0045)	3.69E-02	Ins	0.086(0.0043)	1.72E-82
cg18120259	Chol	-0.011(0.0045)	1.78E-02	Ins	0.087(0.0043)	4.85E-85
cg18181703	Chol	-0.007(0.0045)	1.09E-01	Ins	0.087(0.0043)	1.21E-82
cg18217136	Chol	-0.011(0.0045)	1.37E-02	Ins	0.087(0.0043)	1.64E-84
cg18219562	Chol	-0.010(0.0045)	2.40E-02	Ins	0.087(0.0043)	6.45E-85
cg18513344	Chol	-0.010(0.0045)	2.71E-02	Ins	0.086(0.0043)	3.11E-82
cg18608055	Chol	-0.007(0.0045)	1.42E-01	Ins	0.087(0.0043)	1.57E-82
cg19217955	Chol	-0.011(0.0045)	1.56E-02	Ins	0.086(0.0043)	1.30E-82
cg19373099	Chol	-0.009(0.0045)	5.00E-02	Ins	0.087(0.0043)	3.11E-83
cg19566658	Chol	-0.010(0.0045)	2.52E-02	Ins	0.087(0.0043)	3.81E-84
cg19589396	Chol	-0.010(0.0045)	1.99E-02	Ins	0.087(0.0043)	3.07E-84
cg19695507	Chol	-0.011(0.0045)	1.69E-02	Ins	0.087(0.0043)	2.60E-83
cg19750657	Chol	-0.012(0.0045)	9.55E-03	Ins	0.086(0.0044)	1.75E-81
cg19881557	Chol	-0.009(0.0045)	4.89E-02	Ins	0.087(0.0043)	2.90E-83
cg19998073	Chol	-0.010(0.0045)	1.99E-02	Ins	0.088(0.0043)	1.06E-84

cg21108085	Chol	-0.010(0.0045)	2.35E-02	Ins	0.087(0.0043)	5.17E-85
cg21429551	Chol	-0.010(0.0045)	2.85E-02	Ins	0.088(0.0043)	6.29E-86
cg21486834	Chol	-0.009(0.0045)	4.12E-02	Ins	0.087(0.0043)	3.91E-83
cg22012981	Chol	-0.011(0.0045)	1.85E-02	Ins	0.087(0.0043)	7.14E-84
cg22103219	Chol	-0.010(0.0045)	3.05E-02	Ins	0.087(0.0043)	1.12E-84
cg22488164	Chol	-0.009(0.0045)	5.44E-02	Ins	0.087(0.0043)	3.11E-83
cg22534374	Chol	-0.009(0.0045)	5.60E-02	Ins	0.085(0.0043)	3.35E-81
cg22590032	Chol	-0.009(0.0045)	4.49E-02	Ins	0.087(0.0043)	2.22E-83
cg22695339	Chol	-0.008(0.0045)	7.70E-02	Ins	0.087(0.0043)	4.15E-84
cg22700686	Chol	-0.010(0.0045)	2.64E-02	Ins	0.088(0.0043)	1.12E-84
cg23032421	Chol	-0.011(0.0045)	1.80E-02	Ins	0.088(0.0043)	1.27E-86
cg23232188	Chol	-0.010(0.0045)	2.81E-02	Ins	0.087(0.0043)	5.36E-83
cg24174557	Chol	-0.010(0.0045)	3.04E-02	Ins	0.087(0.0043)	2.85E-83
cg24403644	Chol	-0.010(0.0045)	2.08E-02	Ins	0.088(0.0043)	7.13E-86
cg24469729	Chol	-0.010(0.0045)	3.10E-02	Ins	0.087(0.0043)	4.16E-84
cg24531955	Chol	-0.009(0.0045)	4.73E-02	Ins	0.086(0.0043)	3.60E-83
cg24679890	Chol	-0.010(0.0045)	3.15E-02	Ins	0.088(0.0043)	1.35E-86
cg25001190	Chol	-0.009(0.0045)	3.68E-02	Ins	0.087(0.0043)	1.62E-83
cg25096107	Chol	-0.009(0.0045)	5.79E-02	Ins	0.087(0.0043)	9.62E-83
cg25197194	Chol	-0.010(0.0045)	3.00E-02	Ins	0.087(0.0043)	5.51E-85
cg25217710	Chol	-0.012(0.0045)	7.75E-03	Ins	0.085(0.0043)	6.30E-81
cg25435714	Chol	-0.009(0.0045)	5.84E-02	Ins	0.087(0.0043)	1.87E-84
cg25570328	Chol	-0.011(0.0045)	1.04E-02	Ins	0.089(0.0043)	3.27E-87
cg25649826	Chol	-0.010(0.0045)	2.67E-02	Ins	0.087(0.0043)	5.92E-83
cg26033520	Chol	-0.010(0.0045)	3.43E-02	Ins	0.087(0.0043)	1.56E-83
cg26253134	Chol	-0.010(0.0045)	2.57E-02	Ins	0.087(0.0043)	8.12E-85
cg26357885	Chol	-0.010(0.0045)	1.95E-02	Ins	0.088(0.0043)	1.74E-86
cg26361535	Chol	-0.011(0.0045)	1.88E-02	Ins	0.087(0.0043)	2.23E-83
cg26403843	Chol	-0.009(0.0045)	4.20E-02	Ins	0.087(0.0043)	5.29E-83
cg26542660	Chol	-0.010(0.0045)	3.33E-02	Ins	0.087(0.0043)	4.17E-85
cg26663590	Chol	-0.011(0.0045)	1.53E-02	Ins	0.087(0.0043)	8.71E-84
cg26687842	Chol	-0.010(0.0045)	2.67E-02	Ins	0.088(0.0043)	5.60E-85
cg26804423	Chol	-0.009(0.0045)	3.72E-02	Ins	0.086(0.0043)	8.17E-83
cg26836479	Chol	-0.010(0.0045)	2.65E-02	Ins	0.087(0.0043)	2.29E-85
cg26878209	Chol	-0.009(0.0045)	3.86E-02	Ins	0.087(0.0043)	3.62E-83
cg26894079	Chol	-0.009(0.0045)	3.70E-02	Ins	0.087(0.0043)	4.32E-83
cg26952928	Chol	-0.009(0.0045)	4.40E-02	Ins	0.087(0.0043)	5.18E-84
cg27050612	Chol	-0.010(0.0045)	2.58E-02	Ins	0.088(0.0043)	4.97E-85
cg27087650	Chol	-0.010(0.0045)	3.32E-02	Ins	0.088(0.0043)	5.20E-85
cg27115863	Chol	-0.009(0.0045)	3.82E-02	Ins	0.087(0.0043)	9.14E-84
cg27117792	Chol	-0.010(0.0045)	2.62E-02	Ins	0.088(0.0043)	2.21E-85
cg27184903	Chol	-0.010(0.0045)	2.65E-02	Ins	0.088(0.0043)	8.61E-85
cg27269962	Chol	-0.010(0.0045)	2.67E-02	Ins	0.087(0.0043)	4.49E-84
cg27547344	Chol	-0.009(0.0045)	4.85E-02	Ins	0.087(0.0043)	3.23E-84
cg27614723	Chol	-0.011(0.0045)	1.61E-02	Ins	0.088(0.0043)	6.98E-85
ch.2.304154 74F	Chol	-0.010(0.0045)	3.25E-02	Ins	0.087(0.0043)	1.00E-84
cg00094412	CRP	0.047(0.0044)	6.43E-26	LDL	-0.013(0.0046)	6.01E-03
cg00108715	CRP	0.049(0.0044)	1.80E-27	LDL	-0.012(0.0046)	6.28E-03

cg00138407	CRP	0.047(0.0044)	8.98E-26	LDL	-0.012(0.0046)	7.76E-03
cg00144180	CRP	0.048(0.0044)	7.11E-27	LDL	-0.013(0.0046)	5.56E-03
cg00238353	CRP	0.048(0.0044)	8.76E-27	LDL	-0.013(0.0046)	3.90E-03
cg00244001	CRP	0.047(0.0044)	7.54E-26	LDL	-0.013(0.0046)	5.56E-03
cg00431050	CRP	0.046(0.0045)	5.70E-25	LDL	-0.012(0.0046)	6.65E-03
cg00574958	CRP	0.048(0.0045)	6.86E-26	LDL	-0.015(0.0046)	1.25E-03
cg00634542	CRP	0.046(0.0044)	5.06E-25	LDL	-0.013(0.0046)	5.50E-03
cg00673344	CRP	0.048(0.0044)	8.44E-27	LDL	-0.013(0.0045)	5.10E-03
cg00711896	CRP	0.047(0.0044)	6.98E-26	LDL	-0.013(0.0046)	5.00E-03
cg00863378	CRP	0.048(0.0044)	1.46E-26	LDL	-0.012(0.0046)	1.08E-02
cg00973118	CRP	0.048(0.0044)	8.07E-27	LDL	-0.013(0.0046)	4.21E-03
cg01101459	CRP	0.047(0.0044)	3.13E-25	LDL	-0.012(0.0046)	6.75E-03
cg01243823	CRP	0.047(0.0044)	4.39E-26	LDL	-0.013(0.0046)	3.86E-03
cg01511901	CRP	0.047(0.0044)	1.67E-25	LDL	-0.012(0.0046)	1.00E-02
cg01798813	CRP	0.047(0.0045)	1.33E-25	LDL	-0.011(0.0046)	1.61E-02
cg02119938	CRP	0.048(0.0044)	4.87E-27	LDL	-0.013(0.0046)	5.41E-03
cg02286155	CRP	0.048(0.0045)	4.44E-26	LDL	-0.013(0.0046)	5.97E-03
cg02560388	CRP	0.046(0.0044)	1.09E-24	LDL	-0.012(0.0046)	9.49E-03
cg02650017	CRP	0.041(0.0044)	1.04E-20	LDL	-0.011(0.0046)	1.93E-02
cg02711608	CRP	0.047(0.0044)	5.34E-26	LDL	-0.012(0.0046)	7.47E-03
cg02716826	CRP	0.045(0.0044)	5.12E-24	LDL	-0.011(0.0046)	1.36E-02
cg03050965	CRP	0.048(0.0044)	1.22E-26	LDL	-0.013(0.0046)	4.56E-03
cg03159676	CRP	0.047(0.0044)	6.12E-26	LDL	-0.012(0.0046)	1.10E-02
cg03318904	CRP	0.047(0.0045)	2.07E-25	LDL	-0.013(0.0046)	3.97E-03
cg03327570	CRP	0.048(0.0044)	1.50E-26	LDL	-0.012(0.0046)	1.01E-02
cg03433986	CRP	0.048(0.0044)	9.39E-27	LDL	-0.013(0.0046)	5.12E-03
cg03523676	CRP	0.046(0.0044)	4.25E-25	LDL	-0.012(0.0046)	1.10E-02
cg03725309	CRP	0.048(0.0044)	2.80E-26	LDL	-0.012(0.0046)	7.49E-03
cg03885055	CRP	0.047(0.0044)	3.94E-26	LDL	-0.012(0.0046)	8.54E-03
cg03940776	CRP	0.046(0.0044)	3.39E-25	LDL	-0.012(0.0046)	7.53E-03
cg03957124	CRP	0.045(0.0044)	4.57E-24	LDL	-0.012(0.0046)	8.46E-03
cg04011474	CRP	0.047(0.0044)	4.59E-26	LDL	-0.013(0.0046)	5.88E-03
cg04126866	CRP	0.048(0.0044)	2.10E-26	LDL	-0.012(0.0046)	6.30E-03
cg04232128	CRP	0.047(0.0044)	3.21E-25	LDL	-0.013(0.0046)	5.95E-03
cg04524040	CRP	0.048(0.0044)	2.93E-27	LDL	-0.013(0.0045)	5.09E-03
cg04577162	CRP	0.047(0.0044)	1.13E-25	LDL	-0.013(0.0046)	5.79E-03
cg05063895	CRP	0.048(0.0044)	3.50E-27	LDL	-0.013(0.0045)	3.67E-03
cg05095590	CRP	0.048(0.0044)	2.67E-26	LDL	-0.013(0.0046)	4.94E-03
cg05648472	CRP	0.048(0.0045)	8.64E-27	LDL	-0.012(0.0046)	1.14E-02
cg05720226	CRP	0.048(0.0045)	1.59E-26	LDL	-0.012(0.0046)	7.28E-03
cg05845030	CRP	0.047(0.0044)	5.19E-26	LDL	-0.013(0.0046)	5.18E-03
cg06012428	CRP	0.048(0.0044)	2.74E-26	LDL	-0.012(0.0046)	6.18E-03
cg06164260	CRP	0.045(0.0044)	4.01E-24	LDL	-0.013(0.0046)	4.85E-03
cg06192883	CRP	0.045(0.0045)	3.15E-23	LDL	-0.012(0.0046)	9.87E-03
cg06500161	CRP	0.047(0.0045)	5.40E-25	LDL	-0.006(0.0046)	1.92E-01
cg06559575	CRP	0.048(0.0044)	3.50E-26	LDL	-0.012(0.0046)	6.96E-03
cg06603309	CRP	0.047(0.0045)	8.11E-26	LDL	-0.012(0.0046)	1.00E-02
cg06690548	CRP	0.047(0.0044)	5.46E-26	LDL	-0.012(0.0046)	7.41E-03

cg06898549	CRP	0.048(0.0044)	3.22E-26	LDL	-0.013(0.0046)	3.38E-03
cg06946797	CRP	0.046(0.0045)	1.70E-24	LDL	-0.012(0.0046)	8.35E-03
cg07021906	CRP	0.047(0.0044)	2.48E-25	LDL	-0.013(0.0046)	4.38E-03
cg07037944	CRP	0.047(0.0044)	5.52E-26	LDL	-0.013(0.0046)	6.17E-03
cg07136133	CRP	0.047(0.0044)	1.38E-25	LDL	-0.012(0.0046)	6.92E-03
cg07202479	CRP	0.047(0.0044)	1.14E-25	LDL	-0.013(0.0046)	4.27E-03
cg07471614	CRP	0.049(0.0044)	1.26E-27	LDL	-0.013(0.0046)	4.64E-03
cg07504977	CRP	0.048(0.0044)	2.74E-26	LDL	-0.012(0.0046)	9.74E-03
cg07682160	CRP	0.047(0.0045)	1.46E-25	LDL	-0.012(0.0046)	1.15E-02
cg07728579	CRP	0.047(0.0044)	1.69E-25	LDL	-0.013(0.0046)	6.06E-03
cg07769588	CRP	0.048(0.0045)	1.13E-26	LDL	-0.011(0.0046)	1.53E-02
cg08305942	CRP	0.047(0.0044)	4.38E-26	LDL	-0.012(0.0046)	8.08E-03
cg08309687	CRP	0.049(0.0044)	2.83E-27	LDL	-0.013(0.0046)	5.45E-03
cg08443038	CRP	0.047(0.0044)	6.73E-26	LDL	-0.012(0.0046)	6.28E-03
cg08548559	CRP	0.047(0.0044)	1.15E-25	LDL	-0.013(0.0045)	3.62E-03
cg08648047	CRP	0.047(0.0044)	9.27E-26	LDL	-0.012(0.0046)	8.67E-03
cg08726900	CRP	0.048(0.0045)	8.82E-27	LDL	-0.013(0.0046)	3.43E-03
cg08813944	CRP	0.047(0.0044)	7.23E-26	LDL	-0.013(0.0046)	4.03E-03
cg08857797	CRP	0.047(0.0045)	3.20E-25	LDL	-0.013(0.0046)	3.65E-03
cg09152259	CRP	0.047(0.0044)	1.37E-25	LDL	-0.012(0.0046)	1.07E-02
cg09222732	CRP	0.048(0.0044)	7.56E-27	LDL	-0.012(0.0046)	7.62E-03
cg09315878	CRP	0.048(0.0044)	5.52E-27	LDL	-0.013(0.0046)	5.14E-03
cg09349128	CRP	0.043(0.0045)	3.06E-21	LDL	-0.010(0.0046)	2.68E-02
cg09554443	CRP	0.048(0.0045)	5.15E-26	LDL	-0.013(0.0046)	5.11E-03
cg09613192	CRP	0.047(0.0044)	1.31E-25	LDL	-0.013(0.0046)	5.02E-03
cg09664445	CRP	0.047(0.0045)	1.75E-25	LDL	-0.012(0.0046)	7.11E-03
cg09777883	CRP	0.047(0.0044)	4.31E-26	LDL	-0.013(0.0046)	4.72E-03
cg10179300	CRP	0.048(0.0045)	2.18E-26	LDL	-0.012(0.0046)	7.15E-03
cg10438589	CRP	0.047(0.0044)	9.87E-26	LDL	-0.013(0.0046)	5.70E-03
cg10505902	CRP	0.047(0.0044)	5.98E-26	LDL	-0.012(0.0046)	8.63E-03
cg10513161	CRP	0.047(0.0044)	8.08E-26	LDL	-0.012(0.0046)	8.05E-03
cg10549088	CRP	0.047(0.0045)	6.75E-26	LDL	-0.012(0.0046)	8.43E-03
cg10717869	CRP	0.048(0.0044)	8.91E-27	LDL	-0.013(0.0046)	4.59E-03
cg10734665	CRP	0.047(0.0044)	9.88E-26	LDL	-0.011(0.0046)	1.37E-02
cg10814005	CRP	0.047(0.0044)	1.30E-25	LDL	-0.013(0.0046)	5.79E-03
cg10919522	CRP	0.046(0.0044)	1.12E-24	LDL	-0.012(0.0046)	9.22E-03
cg10922280	CRP	0.046(0.0044)	1.35E-24	LDL	-0.012(0.0046)	8.41E-03
cg10927968	CRP	0.046(0.0044)	4.15E-24	LDL	-0.014(0.0046)	2.38E-03
cg10975897	CRP	0.048(0.0044)	1.05E-26	LDL	-0.013(0.0046)	4.99E-03
cg11024682	CRP	0.047(0.0045)	1.10E-25	LDL	-0.011(0.0046)	1.46E-02
cg11080651	CRP	0.048(0.0044)	1.90E-26	LDL	-0.013(0.0046)	5.56E-03
cg11183227	CRP	0.046(0.0044)	5.49E-25	LDL	-0.013(0.0046)	5.55E-03
cg11202345	CRP	0.048(0.0045)	1.26E-26	LDL	-0.012(0.0046)	7.03E-03
cg11376147	CRP	0.048(0.0044)	2.71E-26	LDL	-0.013(0.0046)	4.52E-03
cg11614585	CRP	0.047(0.0044)	8.86E-26	LDL	-0.013(0.0046)	5.11E-03
cg11650298	CRP	0.047(0.0044)	7.34E-26	LDL	-0.012(0.0046)	1.10E-02
cg11832534	CRP	0.046(0.0044)	4.62E-25	LDL	-0.012(0.0046)	7.92E-03
cg11927233	CRP	0.048(0.0044)	4.02E-27	LDL	-0.013(0.0046)	5.88E-03

cg11969813	CRP	0.047(0.0044)	3.18E-25	LDL	-0.012(0.0046)	7.00E-03
cg12484113	CRP	0.048(0.0045)	5.83E-26	LDL	-0.013(0.0046)	3.54E-03
cg12593793	CRP	0.047(0.0045)	3.59E-25	LDL	-0.013(0.0046)	3.76E-03
cg12992827	CRP	0.045(0.0044)	1.53E-23	LDL	-0.012(0.0046)	1.15E-02
cg13097800	CRP	0.048(0.0044)	7.86E-27	LDL	-0.013(0.0045)	5.09E-03
cg13123009	CRP	0.048(0.0044)	2.35E-26	LDL	-0.013(0.0046)	3.68E-03
cg13274938	CRP	0.047(0.0044)	1.46E-25	LDL	-0.013(0.0046)	5.71E-03
cg13591783	CRP	0.048(0.0044)	8.67E-27	LDL	-0.013(0.0046)	5.80E-03
cg13781414	CRP	0.046(0.0044)	6.38E-25	LDL	-0.012(0.0046)	9.26E-03
cg13922488	CRP	0.048(0.0044)	2.86E-26	LDL	-0.012(0.0046)	6.65E-03
cg14020176	CRP	0.048(0.0044)	6.61E-27	LDL	-0.013(0.0046)	5.48E-03
cg14264316	CRP	0.048(0.0044)	2.95E-26	LDL	-0.012(0.0046)	7.61E-03
cg14476101	CRP	0.047(0.0044)	7.55E-26	LDL	-0.012(0.0046)	6.89E-03
cg15323828	CRP	0.048(0.0044)	2.14E-26	LDL	-0.012(0.0046)	6.83E-03
cg15357118	CRP	0.046(0.0044)	4.95E-25	LDL	-0.012(0.0046)	9.23E-03
cg15681239	CRP	0.047(0.0044)	5.95E-26	LDL	-0.012(0.0046)	7.73E-03
cg15721584	CRP	0.047(0.0045)	3.81E-25	LDL	-0.012(0.0046)	8.61E-03
cg16163382	CRP	0.046(0.0044)	1.54E-24	LDL	-0.012(0.0046)	8.69E-03
cg16578636	CRP	0.048(0.0044)	8.77E-27	LDL	-0.012(0.0045)	9.06E-03
cg16594806	CRP	0.046(0.0045)	2.49E-24	LDL	-0.012(0.0046)	1.02E-02
cg16611584	CRP	0.046(0.0044)	1.34E-24	LDL	-0.013(0.0046)	3.42E-03
cg16815882	CRP	0.047(0.0045)	1.14E-25	LDL	-0.012(0.0046)	7.43E-03
cg16846518	CRP	0.048(0.0044)	2.07E-26	LDL	-0.012(0.0046)	7.13E-03
cg17178175	CRP	0.047(0.0044)	1.04E-25	LDL	-0.012(0.0046)	1.14E-02
cg17260706	CRP	0.046(0.0045)	2.33E-24	LDL	-0.013(0.0046)	4.26E-03
cg17501210	CRP	0.045(0.0044)	2.23E-23	LDL	-0.011(0.0046)	1.51E-02
cg17901584	CRP	0.048(0.0045)	1.84E-26	LDL	-0.008(0.0045)	7.20E-02
cg17971578	CRP	0.048(0.0044)	8.99E-27	LDL	-0.012(0.0045)	6.48E-03
cg18098839	CRP	0.046(0.0044)	4.43E-25	LDL	-0.012(0.0046)	9.80E-03
cg18120259	CRP	0.047(0.0044)	2.88E-26	LDL	-0.013(0.0046)	5.45E-03
cg18181703	CRP	0.044(0.0044)	8.14E-23	LDL	-0.010(0.0046)	2.64E-02
cg18217136	CRP	0.047(0.0044)	1.02E-25	LDL	-0.014(0.0046)	3.03E-03
cg18219562	CRP	0.048(0.0044)	8.86E-27	LDL	-0.013(0.0046)	6.00E-03
cg18513344	CRP	0.045(0.0044)	9.14E-24	LDL	-0.012(0.0046)	7.18E-03
cg18608055	CRP	0.043(0.0044)	3.57E-22	LDL	-0.010(0.0046)	2.77E-02
cg19217955	CRP	0.048(0.0044)	4.38E-26	LDL	-0.013(0.0046)	5.24E-03
cg19373099	CRP	0.047(0.0045)	1.06E-25	LDL	-0.012(0.0046)	9.11E-03
cg19566658	CRP	0.048(0.0044)	2.36E-26	LDL	-0.013(0.0046)	4.78E-03
cg19589396	CRP	0.047(0.0044)	1.45E-25	LDL	-0.013(0.0046)	4.86E-03
cg19695507	CRP	0.047(0.0045)	6.43E-26	LDL	-0.014(0.0046)	2.70E-03
cg19750657	CRP	0.046(0.0045)	2.31E-24	LDL	-0.014(0.0046)	2.16E-03
cg19881557	CRP	0.046(0.0044)	2.16E-24	LDL	-0.012(0.0046)	1.05E-02
cg19998073	CRP	0.049(0.0045)	1.66E-27	LDL	-0.013(0.0046)	5.28E-03
cg21108085	CRP	0.048(0.0044)	2.91E-26	LDL	-0.013(0.0046)	5.52E-03
cg21429551	CRP	0.048(0.0044)	7.25E-27	LDL	-0.012(0.0045)	7.11E-03
cg21486834	CRP	0.047(0.0044)	2.85E-25	LDL	-0.012(0.0046)	8.65E-03
cg22012981	CRP	0.047(0.0044)	5.65E-26	LDL	-0.013(0.0046)	4.45E-03
cg22103219	CRP	0.046(0.0044)	2.24E-25	LDL	-0.012(0.0046)	7.11E-03

cg22488164	CRP	0.046(0.0044)	1.81E-24	LDL	-0.012(0.0046)	7.27E-03
cg22534374	CRP	0.046(0.0044)	6.76E-25	LDL	-0.011(0.0046)	1.27E-02
cg22590032	CRP	0.047(0.0044)	2.68E-25	LDL	-0.012(0.0046)	9.32E-03
cg22695339	CRP	0.047(0.0045)	5.65E-26	LDL	-0.011(0.0046)	2.08E-02
cg22700686	CRP	0.047(0.0044)	6.77E-26	LDL	-0.013(0.0046)	5.53E-03
cg23032421	CRP	0.049(0.0044)	2.03E-27	LDL	-0.013(0.0045)	3.54E-03
cg23232188	CRP	0.047(0.0045)	8.77E-26	LDL	-0.013(0.0046)	5.52E-03
cg24174557	CRP	0.046(0.0044)	1.79E-24	LDL	-0.012(0.0046)	8.71E-03
cg24403644	CRP	0.048(0.0044)	1.47E-26	LDL	-0.013(0.0046)	4.86E-03
cg24469729	CRP	0.047(0.0044)	1.79E-25	LDL	-0.013(0.0046)	6.22E-03
cg24531955	CRP	0.046(0.0044)	1.06E-24	LDL	-0.011(0.0046)	1.20E-02
cg24679890	CRP	0.048(0.0044)	7.32E-27	LDL	-0.012(0.0046)	7.25E-03
cg25001190	CRP	0.048(0.0044)	2.17E-26	LDL	-0.012(0.0046)	9.47E-03
cg25096107	CRP	0.046(0.0045)	5.78E-25	LDL	-0.012(0.0046)	9.50E-03
cg25197194	CRP	0.048(0.0044)	5.03E-27	LDL	-0.012(0.0046)	6.60E-03
cg25217710	CRP	0.047(0.0045)	3.13E-25	LDL	-0.014(0.0046)	1.60E-03
cg25435714	CRP	0.047(0.0044)	5.56E-26	LDL	-0.011(0.0046)	1.52E-02
cg25570328	CRP	0.048(0.0044)	6.67E-27	LDL	-0.014(0.0046)	2.44E-03
cg25649826	CRP	0.047(0.0044)	2.17E-25	LDL	-0.013(0.0046)	4.99E-03
cg26033520	CRP	0.049(0.0045)	2.23E-27	LDL	-0.012(0.0046)	7.05E-03
cg26253134	CRP	0.048(0.0044)	1.48E-26	LDL	-0.012(0.0046)	6.22E-03
cg26357885	CRP	0.048(0.0044)	8.12E-27	LDL	-0.013(0.0045)	4.50E-03
cg26361535	CRP	0.048(0.0045)	5.21E-27	LDL	-0.013(0.0046)	3.67E-03
cg26403843	CRP	0.048(0.0045)	1.73E-26	LDL	-0.011(0.0046)	1.50E-02
cg26542660	CRP	0.047(0.0044)	6.27E-26	LDL	-0.012(0.0046)	7.35E-03
cg26663590	CRP	0.045(0.0044)	6.62E-24	LDL	-0.013(0.0046)	3.40E-03
cg26687842	CRP	0.048(0.0044)	1.27E-26	LDL	-0.013(0.0046)	5.74E-03
cg26804423	CRP	0.046(0.0044)	2.19E-24	LDL	-0.013(0.0046)	5.26E-03
cg26836479	CRP	0.048(0.0044)	5.86E-27	LDL	-0.012(0.0045)	7.07E-03
cg26878209	CRP	0.047(0.0044)	1.47E-25	LDL	-0.012(0.0046)	9.07E-03
cg26894079	CRP	0.047(0.0044)	1.67E-25	LDL	-0.012(0.0046)	7.87E-03
cg26952928	CRP	0.047(0.0044)	3.54E-25	LDL	-0.012(0.0046)	1.06E-02
cg27050612	CRP	0.046(0.0044)	2.90E-25	LDL	-0.012(0.0046)	7.47E-03
cg27087650	CRP	0.046(0.0044)	3.66E-25	LDL	-0.012(0.0046)	7.64E-03
cg27115863	CRP	0.046(0.0044)	3.68E-25	LDL	-0.012(0.0046)	8.50E-03
cg27117792	CRP	0.048(0.0044)	1.30E-26	LDL	-0.013(0.0046)	5.39E-03
cg27184903	CRP	0.047(0.0044)	3.16E-25	LDL	-0.013(0.0046)	4.93E-03
cg27269962	CRP	0.048(0.0044)	1.99E-26	LDL	-0.012(0.0046)	6.70E-03
cg27547344	CRP	0.048(0.0044)	8.87E-27	LDL	-0.012(0.0046)	9.55E-03
cg27614723	CRP	0.049(0.0044)	1.43E-27	LDL	-0.013(0.0046)	3.22E-03
ch.2.304154 74F	CRP	0.048(0.0044)	1.60E-26	LDL	-0.012(0.0046)	7.40E-03
cg00094412	Glc	0.043(0.0043)	3.96E-23	SBP	0.037(0.0041)	3.54E-19
cg00108715	Glc	0.042(0.0043)	1.66E-22	SBP	0.037(0.0041)	2.09E-19
cg00138407	Glc	0.043(0.0043)	7.15E-23	SBP	0.037(0.0041)	7.07E-20
cg00144180	Glc	0.044(0.0043)	1.16E-23	SBP	0.037(0.0041)	1.01E-19
cg00238353	Glc	0.043(0.0043)	7.75E-23	SBP	0.037(0.0041)	2.93E-19
cg00244001	Glc	0.043(0.0043)	5.74E-23	SBP	0.036(0.0041)	5.14E-19
cg00431050	Glc	0.042(0.0043)	4.07E-22	SBP	0.036(0.0041)	4.28E-18

cg00574958	Glc	0.039(0.0043)	1.08E-18	SBP	0.035(0.0041)	3.40E-17
cg00634542	Glc	0.042(0.0043)	3.64E-22	SBP	0.037(0.0041)	4.86E-19
cg00673344	Glc	0.043(0.0043)	2.01E-23	SBP	0.037(0.0040)	1.01E-19
cg00711896	Glc	0.042(0.0043)	3.20E-22	SBP	0.035(0.0041)	5.16E-18
cg00863378	Glc	0.043(0.0043)	4.46E-23	SBP	0.037(0.0041)	8.34E-20
cg00973118	Glc	0.043(0.0043)	1.40E-22	SBP	0.036(0.0041)	2.91E-18
cg01101459	Glc	0.042(0.0043)	7.18E-22	SBP	0.037(0.0041)	3.94E-19
cg01243823	Glc	0.042(0.0043)	2.22E-22	SBP	0.036(0.0041)	7.44E-19
cg01511901	Glc	0.043(0.0043)	1.33E-22	SBP	0.037(0.0041)	8.89E-20
cg01798813	Glc	0.042(0.0043)	3.34E-22	SBP	0.037(0.0041)	4.37E-19
cg02119938	Glc	0.043(0.0043)	3.47E-23	SBP	0.036(0.0041)	1.75E-18
cg02286155	Glc	0.041(0.0043)	3.05E-21	SBP	0.037(0.0041)	1.93E-19
cg02560388	Glc	0.043(0.0043)	1.46E-22	SBP	0.038(0.0041)	3.86E-20
cg02650017	Glc	0.042(0.0043)	1.29E-21	SBP	0.037(0.0041)	6.49E-19
cg02711608	Glc	0.043(0.0043)	9.96E-23	SBP	0.037(0.0040)	3.18E-19
cg02716826	Glc	0.043(0.0043)	6.81E-23	SBP	0.037(0.0041)	8.70E-20
cg03050965	Glc	0.042(0.0043)	2.92E-22	SBP	0.037(0.0041)	4.22E-19
cg03159676	Glc	0.042(0.0043)	1.01E-21	SBP	0.037(0.0041)	3.07E-19
cg03318904	Glc	0.042(0.0043)	1.36E-21	SBP	0.036(0.0041)	8.15E-19
cg03327570	Glc	0.042(0.0043)	2.31E-22	SBP	0.037(0.0041)	1.43E-19
cg03433986	Glc	0.042(0.0043)	1.60E-22	SBP	0.037(0.0041)	1.24E-19
cg03523676	Glc	0.043(0.0043)	1.03E-22	SBP	0.036(0.0041)	5.98E-19
cg03725309	Glc	0.043(0.0043)	1.17E-22	SBP	0.036(0.0040)	2.74E-18
cg03885055	Glc	0.043(0.0043)	9.63E-23	SBP	0.036(0.0041)	5.44E-19
cg03940776	Glc	0.043(0.0043)	5.16E-23	SBP	0.037(0.0041)	1.64E-19
cg03957124	Glc	0.042(0.0043)	5.14E-22	SBP	0.037(0.0041)	3.22E-19
cg04011474	Glc	0.042(0.0043)	2.23E-22	SBP	0.036(0.0041)	6.31E-19
cg04126866	Glc	0.043(0.0043)	4.40E-23	SBP	0.037(0.0041)	1.22E-19
cg04232128	Glc	0.043(0.0043)	3.12E-23	SBP	0.037(0.0041)	1.05E-19
cg04524040	Glc	0.043(0.0043)	3.99E-23	SBP	0.037(0.0040)	1.17E-19
cg04577162	Glc	0.043(0.0043)	9.53E-23	SBP	0.036(0.0041)	2.19E-18
cg05063895	Glc	0.043(0.0043)	5.44E-23	SBP	0.037(0.0041)	1.70E-19
cg05095590	Glc	0.043(0.0043)	9.74E-23	SBP	0.037(0.0041)	2.33E-19
cg05648472	Glc	0.042(0.0043)	5.29E-22	SBP	0.037(0.0041)	3.87E-19
cg05720226	Glc	0.043(0.0043)	4.19E-23	SBP	0.037(0.0041)	9.89E-20
cg05845030	Glc	0.042(0.0043)	5.83E-22	SBP	0.036(0.0041)	6.47E-19
cg06012428	Glc	0.043(0.0043)	2.02E-23	SBP	0.037(0.0041)	3.20E-19
cg06164260	Glc	0.042(0.0043)	9.45E-22	SBP	0.037(0.0041)	4.13E-19
cg06192883	Glc	0.041(0.0043)	3.78E-21	SBP	0.035(0.0041)	7.24E-18
cg06500161	Glc	0.039(0.0044)	1.64E-18	SBP	0.036(0.0041)	9.62E-18
cg06559575	Glc	0.042(0.0043)	3.48E-22	SBP	0.037(0.0041)	2.47E-19
cg06603309	Glc	0.044(0.0043)	1.13E-23	SBP	0.037(0.0041)	1.34E-19
cg06690548	Glc	0.042(0.0043)	1.85E-22	SBP	0.036(0.0040)	2.13E-18
cg06898549	Glc	0.042(0.0043)	1.85E-22	SBP	0.036(0.0041)	5.91E-19
cg06946797	Glc	0.042(0.0043)	3.79E-22	SBP	0.036(0.0041)	1.03E-18
cg07021906	Glc	0.043(0.0043)	8.67E-23	SBP	0.036(0.0041)	8.17E-19
cg07037944	Glc	0.042(0.0043)	4.07E-22	SBP	0.036(0.0041)	4.99E-19
cg07136133	Glc	0.043(0.0043)	1.28E-22	SBP	0.037(0.0041)	3.13E-19

cg07202479	Glc	0.042(0.0043)	4.50E-22	SBP	0.036(0.0041)	4.93E-19
cg07471614	Glc	0.043(0.0043)	5.58E-23	SBP	0.036(0.0041)	2.88E-18
cg07504977	Glc	0.042(0.0043)	3.24E-22	SBP	0.036(0.0041)	1.14E-18
cg07682160	Glc	0.043(0.0043)	8.14E-23	SBP	0.038(0.0041)	5.06E-20
cg07728579	Glc	0.042(0.0043)	5.42E-22	SBP	0.036(0.0041)	1.26E-18
cg07769588	Glc	0.042(0.0043)	8.26E-22	SBP	0.037(0.0041)	4.52E-19
cg08305942	Glc	0.043(0.0043)	4.76E-23	SBP	0.037(0.0041)	9.47E-20
cg08309687	Glc	0.042(0.0043)	1.14E-21	SBP	0.036(0.0041)	6.65E-19
cg08443038	Glc	0.043(0.0043)	8.30E-23	SBP	0.036(0.0041)	4.87E-19
cg08548559	Glc	0.043(0.0043)	1.15E-22	SBP	0.036(0.0040)	4.18E-19
cg08648047	Glc	0.044(0.0043)	1.46E-23	SBP	0.037(0.0041)	4.95E-19
cg08726900	Glc	0.042(0.0043)	3.77E-22	SBP	0.036(0.0041)	6.47E-19
cg08813944	Glc	0.042(0.0043)	2.58E-22	SBP	0.037(0.0041)	3.32E-19
cg08857797	Glc	0.041(0.0043)	4.37E-21	SBP	0.035(0.0041)	3.49E-17
cg09152259	Glc	0.043(0.0043)	1.69E-22	SBP	0.037(0.0041)	2.93E-19
cg09222732	Glc	0.043(0.0043)	8.45E-23	SBP	0.037(0.0041)	1.67E-19
cg09315878	Glc	0.044(0.0043)	1.40E-23	SBP	0.037(0.0041)	1.65E-19
cg09349128	Glc	0.042(0.0044)	3.36E-21	SBP	0.036(0.0041)	1.84E-18
cg09554443	Glc	0.043(0.0043)	5.07E-23	SBP	0.038(0.0041)	4.26E-20
cg09613192	Glc	0.042(0.0043)	2.87E-22	SBP	0.037(0.0041)	2.34E-19
cg09664445	Glc	0.043(0.0043)	2.20E-22	SBP	0.036(0.0041)	5.87E-18
cg09777883	Glc	0.042(0.0043)	2.98E-22	SBP	0.037(0.0041)	3.77E-19
cg10179300	Glc	0.042(0.0043)	4.33E-22	SBP	0.037(0.0041)	5.15E-19
cg10438589	Glc	0.042(0.0043)	2.53E-22	SBP	0.038(0.0041)	3.39E-20
cg10505902	Glc	0.042(0.0043)	2.38E-22	SBP	0.037(0.0041)	3.95E-19
cg10513161	Glc	0.043(0.0043)	3.95E-23	SBP	0.037(0.0041)	3.74E-19
cg10549088	Glc	0.043(0.0043)	3.00E-23	SBP	0.037(0.0041)	4.60E-19
cg10717869	Glc	0.043(0.0043)	2.11E-23	SBP	0.037(0.0041)	9.58E-20
cg10734665	Glc	0.042(0.0043)	3.76E-22	SBP	0.037(0.0041)	8.92E-20
cg10814005	Glc	0.043(0.0043)	4.55E-23	SBP	0.037(0.0041)	4.11E-19
cg10919522	Glc	0.042(0.0043)	2.53E-22	SBP	0.037(0.0041)	4.24E-19
cg10922280	Glc	0.043(0.0043)	6.92E-23	SBP	0.036(0.0041)	1.19E-18
cg10927968	Glc	0.042(0.0043)	7.15E-22	SBP	0.036(0.0041)	1.61E-18
cg10975897	Glc	0.043(0.0043)	1.46E-22	SBP	0.036(0.0041)	1.92E-18
cg11024682	Glc	0.039(0.0043)	2.22E-19	SBP	0.034(0.0041)	6.02E-17
cg11080651	Glc	0.042(0.0043)	2.45E-22	SBP	0.036(0.0041)	5.26E-19
cg11183227	Glc	0.042(0.0043)	2.83E-22	SBP	0.035(0.0041)	1.37E-17
cg11202345	Glc	0.043(0.0043)	1.41E-22	SBP	0.036(0.0041)	1.22E-18
cg11376147	Glc	0.042(0.0043)	3.98E-22	SBP	0.036(0.0041)	1.51E-18
cg11614585	Glc	0.042(0.0043)	6.83E-22	SBP	0.037(0.0041)	2.01E-19
cg11650298	Glc	0.043(0.0043)	1.37E-22	SBP	0.038(0.0041)	4.19E-20
cg11832534	Glc	0.043(0.0043)	8.08E-23	SBP	0.037(0.0041)	4.37E-19
cg11927233	Glc	0.043(0.0043)	8.49E-23	SBP	0.037(0.0041)	2.65E-19
cg11969813	Glc	0.042(0.0043)	8.87E-22	SBP	0.037(0.0041)	3.88E-19
cg12484113	Glc	0.041(0.0043)	1.06E-20	SBP	0.036(0.0041)	1.02E-18
cg12593793	Glc	0.042(0.0043)	4.33E-22	SBP	0.036(0.0041)	3.37E-18
cg12992827	Glc	0.044(0.0043)	8.03E-24	SBP	0.037(0.0041)	1.87E-19
cg13097800	Glc	0.043(0.0043)	7.18E-23	SBP	0.037(0.0040)	3.03E-19

cg13123009	Glc	0.042(0.0043)	2.92E-22	SBP	0.036(0.0041)	8.91E-19
cg13274938	Glc	0.042(0.0043)	2.67E-22	SBP	0.036(0.0041)	7.37E-19
cg13591783	Glc	0.043(0.0043)	3.77E-23	SBP	0.037(0.0041)	8.69E-20
cg13781414	Glc	0.043(0.0043)	9.61E-23	SBP	0.037(0.0041)	7.92E-20
cg13922488	Glc	0.043(0.0043)	5.91E-23	SBP	0.036(0.0041)	5.22E-19
cg14020176	Glc	0.042(0.0043)	5.45E-22	SBP	0.036(0.0041)	1.48E-18
cg14264316	Glc	0.043(0.0043)	3.68E-23	SBP	0.037(0.0041)	7.77E-20
cg14476101	Glc	0.042(0.0043)	2.23E-22	SBP	0.036(0.0041)	1.73E-18
cg15323828	Glc	0.043(0.0043)	4.92E-23	SBP	0.036(0.0041)	7.13E-19
cg15357118	Glc	0.043(0.0043)	1.27E-22	SBP	0.036(0.0041)	1.49E-18
cg15681239	Glc	0.042(0.0043)	1.75E-22	SBP	0.036(0.0041)	9.76E-19
cg15721584	Glc	0.042(0.0043)	1.34E-21	SBP	0.037(0.0041)	7.83E-19
cg16163382	Glc	0.043(0.0043)	3.23E-23	SBP	0.037(0.0041)	2.69E-19
cg16578636	Glc	0.043(0.0043)	2.79E-23	SBP	0.037(0.0040)	1.62E-19
cg16594806	Glc	0.043(0.0044)	5.14E-23	SBP	0.036(0.0041)	1.02E-18
cg16611584	Glc	0.042(0.0043)	3.69E-22	SBP	0.036(0.0041)	7.35E-19
cg16815882	Glc	0.043(0.0043)	1.13E-22	SBP	0.037(0.0041)	5.02E-19
cg16846518	Glc	0.044(0.0043)	8.99E-24	SBP	0.037(0.0041)	1.19E-19
cg17178175	Glc	0.043(0.0043)	4.95E-23	SBP	0.037(0.0041)	3.54E-19
cg17260706	Glc	0.042(0.0043)	3.03E-22	SBP	0.038(0.0041)	6.01E-20
cg17501210	Glc	0.044(0.0043)	2.92E-23	SBP	0.037(0.0041)	7.73E-19
cg17901584	Glc	0.043(0.0043)	1.17E-22	SBP	0.038(0.0041)	2.11E-20
cg17971578	Glc	0.043(0.0043)	5.20E-23	SBP	0.037(0.0040)	1.53E-19
cg18098839	Glc	0.043(0.0043)	3.22E-23	SBP	0.036(0.0041)	5.72E-19
cg18120259	Glc	0.042(0.0043)	1.25E-22	SBP	0.036(0.0040)	7.38E-19
cg18181703	Glc	0.043(0.0043)	1.56E-22	SBP	0.037(0.0041)	1.95E-19
cg18217136	Glc	0.043(0.0043)	1.33E-22	SBP	0.036(0.0041)	1.08E-18
cg18219562	Glc	0.042(0.0043)	2.09E-22	SBP	0.036(0.0041)	6.20E-19
cg18513344	Glc	0.043(0.0043)	1.60E-22	SBP	0.036(0.0041)	8.17E-19
cg18608055	Glc	0.043(0.0043)	5.05E-23	SBP	0.038(0.0041)	5.04E-20
cg19217955	Glc	0.042(0.0043)	3.32E-22	SBP	0.037(0.0041)	4.94E-19
cg19373099	Glc	0.043(0.0043)	2.02E-22	SBP	0.036(0.0041)	1.15E-18
cg19566658	Glc	0.043(0.0043)	3.47E-23	SBP	0.037(0.0041)	2.34E-19
cg19589396	Glc	0.042(0.0043)	2.06E-22	SBP	0.036(0.0041)	4.91E-19
cg19695507	Glc	0.042(0.0043)	7.00E-22	SBP	0.035(0.0041)	6.84E-18
cg19750657	Glc	0.041(0.0044)	2.28E-20	SBP	0.036(0.0041)	4.41E-18
cg19881557	Glc	0.042(0.0043)	3.89E-22	SBP	0.037(0.0041)	5.45E-19
cg19998073	Glc	0.043(0.0043)	1.80E-22	SBP	0.036(0.0041)	2.15E-18
cg21108085	Glc	0.044(0.0043)	7.13E-24	SBP	0.037(0.0041)	1.13E-19
cg21429551	Glc	0.043(0.0043)	9.29E-23	SBP	0.037(0.0040)	3.31E-19
cg21486834	Glc	0.044(0.0043)	8.55E-24	SBP	0.036(0.0041)	5.98E-19
cg22012981	Glc	0.043(0.0043)	1.23E-22	SBP	0.037(0.0041)	1.81E-19
cg22103219	Glc	0.042(0.0043)	1.32E-22	SBP	0.036(0.0040)	8.14E-19
cg22488164	Glc	0.043(0.0043)	4.63E-23	SBP	0.036(0.0041)	1.28E-18
cg22534374	Glc	0.043(0.0043)	1.09E-22	SBP	0.037(0.0041)	3.11E-19
cg22590032	Glc	0.042(0.0043)	6.14E-22	SBP	0.036(0.0041)	2.31E-18
cg22695339	Glc	0.043(0.0043)	3.57E-23	SBP	0.037(0.0041)	3.27E-19
cg22700686	Glc	0.042(0.0043)	2.03E-22	SBP	0.037(0.0041)	8.22E-20

cg23032421	Glc	0.043(0.0043)	6.03E-23	SBP	0.037(0.0040)	9.19E-20
cg23232188	Glc	0.041(0.0043)	2.80E-21	SBP	0.036(0.0041)	1.71E-18
cg24174557	Glc	0.042(0.0043)	8.51E-22	SBP	0.036(0.0041)	1.22E-18
cg24403644	Glc	0.043(0.0043)	2.31E-23	SBP	0.037(0.0041)	2.05E-19
cg24469729	Glc	0.042(0.0043)	7.36E-22	SBP	0.037(0.0041)	1.29E-19
cg24531955	Glc	0.043(0.0043)	6.04E-23	SBP	0.036(0.0041)	5.55E-19
cg24679890	Glc	0.044(0.0043)	9.33E-24	SBP	0.037(0.0041)	5.94E-20
cg25001190	Glc	0.042(0.0043)	1.98E-22	SBP	0.037(0.0041)	1.34E-19
cg25096107	Glc	0.043(0.0043)	3.24E-23	SBP	0.037(0.0041)	2.71E-19
cg25197194	Glc	0.043(0.0043)	7.03E-23	SBP	0.037(0.0041)	1.74E-19
cg25217710	Glc	0.040(0.0043)	1.55E-20	SBP	0.036(0.0041)	9.64E-19
cg25435714	Glc	0.042(0.0043)	1.28E-21	SBP	0.036(0.0041)	9.23E-19
cg25570328	Glc	0.042(0.0043)	2.25E-22	SBP	0.036(0.0041)	1.24E-18
cg25649826	Glc	0.043(0.0043)	7.49E-23	SBP	0.038(0.0041)	4.65E-20
cg26033520	Glc	0.043(0.0043)	1.40E-22	SBP	0.036(0.0041)	6.17E-19
cg26253134	Glc	0.042(0.0043)	4.07E-22	SBP	0.037(0.0041)	2.07E-19
cg26357885	Glc	0.043(0.0043)	3.67E-23	SBP	0.037(0.0040)	2.02E-19
cg26361535	Glc	0.042(0.0043)	2.42E-22	SBP	0.037(0.0041)	3.52E-19
cg26403843	Glc	0.042(0.0043)	3.87E-22	SBP	0.037(0.0041)	4.01E-19
cg26542660	Glc	0.043(0.0043)	3.10E-23	SBP	0.037(0.0041)	1.08E-19
cg26663590	Glc	0.042(0.0043)	4.55E-22	SBP	0.036(0.0041)	3.14E-18
cg26687842	Glc	0.043(0.0043)	1.40E-22	SBP	0.037(0.0041)	3.67E-19
cg26804423	Glc	0.043(0.0043)	1.57E-22	SBP	0.036(0.0041)	5.85E-19
cg26836479	Glc	0.042(0.0043)	1.16E-22	SBP	0.037(0.0040)	2.59E-19
cg26878209	Glc	0.043(0.0043)	5.26E-23	SBP	0.037(0.0041)	1.20E-19
cg26894079	Glc	0.043(0.0043)	1.27E-22	SBP	0.037(0.0041)	2.34E-19
cg26952928	Glc	0.043(0.0043)	1.42E-22	SBP	0.036(0.0041)	6.27E-19
cg27050612	Glc	0.043(0.0043)	6.35E-23	SBP	0.038(0.0041)	3.76E-20
cg27087650	Glc	0.043(0.0043)	1.61E-23	SBP	0.038(0.0041)	1.64E-20
cg27115863	Glc	0.043(0.0043)	5.34E-23	SBP	0.037(0.0041)	2.74E-19
cg27117792	Glc	0.043(0.0043)	2.80E-23	SBP	0.037(0.0041)	1.13E-19
cg27184903	Glc	0.043(0.0043)	1.28E-22	SBP	0.036(0.0041)	2.61E-18
cg27269962	Glc	0.043(0.0043)	8.55E-23	SBP	0.036(0.0041)	5.90E-19
cg27547344	Glc	0.042(0.0043)	2.86E-22	SBP	0.037(0.0041)	3.38E-19
cg27614723	Glc	0.043(0.0043)	4.43E-23	SBP	0.038(0.0041)	3.26E-20
ch.2.30415474F	Glc	0.043(0.0043)	3.11E-23	SBP	0.037(0.0041)	2.70E-19
cg00094412	HbA1c	0.045(0.0043)	1.89E-25	TG	0.027(0.0044)	1.19E-09
cg00108715	HbA1c	0.045(0.0043)	2.85E-25	TG	0.027(0.0044)	8.08E-10
cg00138407	HbA1c	0.045(0.0043)	2.72E-25	TG	0.027(0.0044)	8.15E-10
cg00144180	HbA1c	0.046(0.0043)	2.71E-26	TG	0.026(0.0044)	2.39E-09
cg00238353	HbA1c	0.044(0.0043)	1.16E-24	TG	0.026(0.0044)	2.58E-09
cg00244001	HbA1c	0.045(0.0043)	2.12E-25	TG	0.026(0.0044)	3.18E-09
cg00431050	HbA	0.044(0.0043)	1.41E-24	TG	0.025(0.0044)	1.07E-08

	HbA1c					
cg00574958	HbA1c	0.043(0.0043)	2.07E-23	TG	0.018(0.0044)	2.27E-05
cg00634542	HbA1c	0.044(0.0043)	8.71E-25	TG	0.027(0.0044)	1.06E-09
cg00673344	HbA1c	0.045(0.0043)	4.09E-26	TG	0.027(0.0044)	1.20E-09
cg00711896	HbA1c	0.045(0.0043)	1.66E-25	TG	0.026(0.0044)	2.63E-09
cg00863378	HbA1c	0.045(0.0043)	2.72E-25	TG	0.026(0.0044)	2.40E-09
cg00973118	HbA1c	0.045(0.0043)	2.31E-25	TG	0.027(0.0044)	8.12E-10
cg01101459	HbA1c	0.044(0.0043)	8.18E-25	TG	0.027(0.0044)	1.36E-09
cg01243823	HbA1c	0.045(0.0043)	1.91E-25	TG	0.026(0.0044)	3.96E-09
cg01511901	HbA1c	0.044(0.0043)	5.25E-25	TG	0.028(0.0044)	4.17E-10
cg01798813	HbA1c	0.044(0.0043)	1.40E-24	TG	0.028(0.0044)	3.63E-10
cg02119938	HbA1c	0.045(0.0043)	4.00E-25	TG	0.027(0.0044)	7.67E-10
cg02286155	HbA1c	0.044(0.0043)	4.47E-24	TG	0.026(0.0044)	2.88E-09
cg02560388	HbA1c	0.044(0.0043)	3.35E-24	TG	0.026(0.0044)	3.39E-09
cg02650017	HbA1c	0.044(0.0043)	4.66E-24	TG	0.027(0.0044)	1.51E-09
cg02711608	HbA1c	0.045(0.0042)	2.16E-25	TG	0.026(0.0044)	4.30E-09
cg02716826	HbA1c	0.045(0.0043)	5.09E-25	TG	0.026(0.0044)	3.02E-09
cg03050965	HbA1c	0.045(0.0043)	4.45E-25	TG	0.026(0.0044)	2.41E-09
cg03159676	HbA1c	0.045(0.0043)	3.23E-25	TG	0.026(0.0044)	3.23E-09
cg03318904	HbA1c	0.044(0.0043)	1.52E-24	TG	0.025(0.0044)	1.11E-08
cg03327570	HbA1c	0.045(0.0043)	1.77E-25	TG	0.026(0.0044)	3.97E-09
cg03433986	HbA1c	0.045(0.0042)	2.36E-25	TG	0.026(0.0044)	1.96E-09
cg03523676	HbA1c	0.044(0.0043)	2.11E-24	TG	0.028(0.0044)	3.13E-10
cg03725309	HbA1c	0.044(0.0043)	5.70E-25	TG	0.025(0.0044)	1.63E-08
cg03885055	HbA1c	0.045(0.0043)	1.00E-25	TG	0.027(0.0044)	1.13E-09
cg03940776	HbA1c	0.045(0.0043)	3.50E-25	TG	0.027(0.0044)	6.68E-10
cg03957124	HbA1c	0.044(0.0043)	3.04E-24	TG	0.027(0.0044)	2.29E-09
cg04011474	HbA1c	0.045(0.0043)	2.19E-25	TG	0.026(0.0044)	2.20E-09

	HbA1c					
cg04126866	HbA1c	0.045(0.0043)	7.86E-26	TG	0.026(0.0044)	2.36E-09
cg04232128	HbA1c	0.044(0.0043)	5.44E-25	TG	0.027(0.0044)	1.62E-09
cg04524040	HbA1c	0.045(0.0042)	5.03E-26	TG	0.027(0.0044)	1.65E-09
cg04577162	HbA1c	0.044(0.0043)	7.92E-25	TG	0.027(0.0044)	1.58E-09
cg05063895	HbA1c	0.046(0.0043)	3.14E-26	TG	0.026(0.0044)	1.88E-09
cg05095590	HbA1c	0.045(0.0042)	2.29E-25	TG	0.026(0.0044)	2.04E-09
cg05648472	HbA1c	0.043(0.0043)	7.93E-24	TG	0.027(0.0044)	1.83E-09
cg05720226	HbA1c	0.046(0.0043)	1.46E-26	TG	0.026(0.0044)	4.83E-09
cg05845030	HbA1c	0.044(0.0043)	1.39E-24	TG	0.027(0.0044)	1.99E-09
cg06012428	HbA1c	0.045(0.0043)	6.61E-26	TG	0.026(0.0044)	3.12E-09
cg06164260	HbA1c	0.044(0.0043)	4.06E-24	TG	0.026(0.0044)	7.53E-09
cg06192883	HbA1c	0.044(0.0043)	7.72E-24	TG	0.026(0.0044)	8.41E-09
cg06500161	HbA1c	0.041(0.0043)	1.82E-21	TG	0.018(0.0044)	2.85E-05
cg06559575	HbA1c	0.044(0.0043)	5.41E-25	TG	0.026(0.0044)	6.48E-09
cg06603309	HbA1c	0.045(0.0043)	1.78E-25	TG	0.027(0.0044)	1.17E-09
cg06690548	HbA1c	0.045(0.0043)	2.10E-25	TG	0.024(0.0044)	2.45E-08
cg06898549	HbA1c	0.044(0.0043)	6.31E-25	TG	0.026(0.0044)	2.02E-09
cg06946797	HbA1c	0.044(0.0043)	2.18E-24	TG	0.026(0.0044)	3.22E-09
cg07021906	HbA1c	0.045(0.0043)	3.04E-25	TG	0.026(0.0044)	5.49E-09
cg07037944	HbA1c	0.045(0.0043)	1.40E-25	TG	0.027(0.0044)	1.77E-09
cg07136133	HbA1c	0.045(0.0043)	1.71E-25	TG	0.026(0.0044)	3.24E-09
cg07202479	HbA1c	0.045(0.0043)	2.27E-25	TG	0.026(0.0044)	6.21E-09
cg07471614	HbA1c	0.045(0.0043)	2.73E-25	TG	0.026(0.0044)	4.61E-09
cg07504977	HbA1c	0.045(0.0043)	2.25E-25	TG	0.024(0.0044)	3.73E-08
cg07682160	HbA1c	0.045(0.0043)	3.83E-25	TG	0.026(0.0044)	8.87E-09
cg07728579	HbA1c	0.045(0.0043)	1.59E-25	TG	0.025(0.0044)	8.24E-09
cg07769588	HbA1c	0.044(0.0043)	3.11E-24	TG	0.027(0.0044)	1.86E-09

	HbA1c					
cg08305942	HbA1c	0.045(0.0043)	7.24E-26	TG	0.027(0.0044)	8.73E-10
cg08309687	HbA1c	0.043(0.0043)	6.86E-24	TG	0.025(0.0044)	8.14E-09
cg08443038	HbA1c	0.046(0.0043)	2.99E-26	TG	0.026(0.0044)	2.18E-09
cg08548559	HbA1c	0.045(0.0043)	1.15E-25	TG	0.026(0.0044)	2.41E-09
cg08648047	HbA1c	0.045(0.0043)	9.90E-26	TG	0.026(0.0044)	5.98E-09
cg08726900	HbA1c	0.043(0.0043)	1.09E-23	TG	0.025(0.0044)	1.06E-08
cg08813944	HbA1c	0.045(0.0043)	4.41E-25	TG	0.028(0.0044)	3.00E-10
cg08857797	HbA1c	0.043(0.0043)	7.82E-24	TG	0.025(0.0044)	2.54E-08
cg09152259	HbA1c	0.044(0.0043)	8.56E-25	TG	0.027(0.0044)	1.82E-09
cg09222732	HbA1c	0.045(0.0043)	7.35E-26	TG	0.026(0.0044)	2.58E-09
cg09315878	HbA1c	0.046(0.0043)	4.18E-26	TG	0.026(0.0044)	2.94E-09
cg09349128	HbA1c	0.042(0.0043)	2.06E-22	TG	0.028(0.0045)	4.70E-10
cg09554443	HbA1c	0.045(0.0043)	3.15E-25	TG	0.025(0.0044)	1.46E-08
cg09613192	HbA1c	0.044(0.0043)	3.00E-24	TG	0.026(0.0044)	2.76E-09
cg09664445	HbA1c	0.044(0.0043)	1.51E-24	TG	0.027(0.0044)	1.26E-09
cg09777883	HbA1c	0.044(0.0042)	1.03E-24	TG	0.026(0.0044)	3.42E-09
cg10179300	HbA1c	0.045(0.0043)	2.76E-25	TG	0.026(0.0044)	4.40E-09
cg10438589	HbA1c	0.045(0.0043)	3.22E-25	TG	0.027(0.0044)	9.30E-10
cg10505902	HbA1c	0.045(0.0043)	1.37E-25	TG	0.027(0.0044)	1.06E-09
cg10513161	HbA1c	0.045(0.0043)	1.12E-25	TG	0.027(0.0044)	7.66E-10
cg10549088	HbA1c	0.045(0.0043)	2.30E-25	TG	0.027(0.0044)	1.30E-09
cg10717869	HbA1c	0.045(0.0043)	6.80E-26	TG	0.026(0.0044)	2.64E-09
cg10734665	HbA1c	0.045(0.0043)	2.99E-25	TG	0.027(0.0044)	6.67E-10
cg10814005	HbA1c	0.044(0.0043)	4.92E-25	TG	0.027(0.0044)	1.78E-09
cg10919522	HbA1c	0.044(0.0043)	1.04E-24	TG	0.026(0.0044)	2.71E-09
cg10922280	HbA1c	0.045(0.0043)	5.50E-25	TG	0.025(0.0044)	9.17E-09
cg10927968	HbA1c	0.043(0.0043)	1.47E-23	TG	0.026(0.0044)	6.64E-09

	HbA1c					
cg10975897	HbA1c	0.044(0.0043)	6.90E-25	TG	0.025(0.0044)	8.35E-09
cg11024682	HbA1c	0.043(0.0043)	1.90E-23	TG	0.021(0.0044)	2.00E-06
cg11080651	HbA1c	0.044(0.0042)	1.20E-24	TG	0.026(0.0044)	2.18E-09
cg11183227	HbA1c	0.044(0.0043)	1.22E-24	TG	0.027(0.0044)	7.11E-10
cg11202345	HbA1c	0.043(0.0043)	6.20E-24	TG	0.026(0.0044)	2.90E-09
cg11376147	HbA1c	0.045(0.0043)	2.90E-25	TG	0.025(0.0044)	1.10E-08
cg11614585	HbA1c	0.044(0.0043)	1.69E-24	TG	0.025(0.0044)	1.90E-08
cg11650298	HbA1c	0.044(0.0043)	5.41E-25	TG	0.026(0.0044)	4.23E-09
cg11832534	HbA1c	0.045(0.0043)	3.55E-25	TG	0.027(0.0044)	1.14E-09
cg11927233	HbA1c	0.045(0.0043)	1.26E-25	TG	0.026(0.0044)	3.70E-09
cg11969813	HbA1c	0.044(0.0043)	1.42E-24	TG	0.026(0.0044)	4.28E-09
cg12484113	HbA1c	0.045(0.0043)	6.97E-25	TG	0.025(0.0044)	1.43E-08
cg12593793	HbA1c	0.044(0.0043)	3.65E-24	TG	0.026(0.0044)	8.61E-09
cg12992827	HbA1c	0.044(0.0043)	7.12E-24	TG	0.027(0.0044)	9.90E-10
cg13097800	HbA1c	0.045(0.0043)	6.40E-26	TG	0.027(0.0044)	9.27E-10
cg13123009	HbA1c	0.044(0.0043)	1.02E-24	TG	0.026(0.0044)	5.70E-09
cg13274938	HbA1c	0.044(0.0043)	6.12E-25	TG	0.027(0.0044)	1.29E-09
cg13591783	HbA1c	0.045(0.0043)	1.30E-25	TG	0.027(0.0044)	1.51E-09
cg13781414	HbA1c	0.045(0.0043)	2.05E-25	TG	0.026(0.0044)	2.48E-09
cg13922488	HbA1c	0.045(0.0043)	2.83E-25	TG	0.027(0.0044)	1.01E-09
cg14020176	HbA1c	0.045(0.0043)	4.69E-25	TG	0.026(0.0044)	4.12E-09
cg14264316	HbA1c	0.045(0.0043)	7.77E-26	TG	0.027(0.0044)	8.67E-10
cg14476101	HbA1c	0.045(0.0043)	1.49E-25	TG	0.024(0.0044)	4.73E-08
cg15323828	HbA1c	0.046(0.0043)	3.60E-26	TG	0.027(0.0044)	1.92E-09
cg15357118	HbA1c	0.044(0.0043)	8.49E-25	TG	0.026(0.0044)	3.66E-09
cg15681239	HbA1c	0.044(0.0043)	5.21E-25	TG	0.028(0.0044)	5.00E-10
cg15721584	HbA1c	0.043(0.0043)	1.12E-23	TG	0.026(0.0044)	5.88E-09

	HbA1c					
cg16163382	HbA1c	0.045(0.0043)	2.76E-25	TG	0.027(0.0044)	1.17E-09
cg16578636	HbA1c	0.045(0.0042)	4.50E-26	TG	0.027(0.0044)	9.43E-10
cg16594806	HbA1c	0.045(0.0043)	2.86E-25	TG	0.027(0.0044)	2.51E-09
cg16611584	HbA1c	0.045(0.0043)	3.01E-25	TG	0.026(0.0044)	3.42E-09
cg16815882	HbA1c	0.045(0.0043)	2.27E-25	TG	0.026(0.0044)	2.43E-09
cg16846518	HbA1c	0.045(0.0043)	4.54E-26	TG	0.027(0.0044)	8.88E-10
cg17178175	HbA1c	0.045(0.0043)	1.23E-25	TG	0.027(0.0044)	8.24E-10
cg17260706	HbA1c	0.044(0.0043)	9.09E-25	TG	0.025(0.0044)	9.97E-09
cg17501210	HbA1c	0.043(0.0043)	2.80E-23	TG	0.027(0.0044)	1.18E-09
cg17901584	HbA1c	0.042(0.0043)	6.25E-23	TG	0.025(0.0044)	1.64E-08
cg17971578	HbA1c	0.045(0.0042)	8.95E-26	TG	0.027(0.0044)	1.65E-09
cg18098839	HbA1c	0.044(0.0043)	1.17E-24	TG	0.027(0.0044)	1.96E-09
cg18120259	HbA1c	0.045(0.0043)	1.60E-25	TG	0.025(0.0044)	7.54E-09
cg18181703	HbA1c	0.044(0.0043)	1.98E-24	TG	0.027(0.0044)	1.86E-09
cg18217136	HbA1c	0.044(0.0043)	7.13E-25	TG	0.026(0.0044)	2.10E-09
cg18219562	HbA1c	0.045(0.0043)	1.99E-25	TG	0.026(0.0044)	3.68E-09
cg18513344	HbA1c	0.045(0.0043)	3.11E-25	TG	0.025(0.0044)	1.40E-08
cg18608055	HbA1c	0.043(0.0043)	1.14E-23	TG	0.027(0.0044)	7.97E-10
cg19217955	HbA1c	0.045(0.0043)	5.20E-25	TG	0.025(0.0044)	1.17E-08
cg19373099	HbA1c	0.045(0.0043)	2.96E-25	TG	0.027(0.0044)	1.00E-09
cg19566658	HbA1c	0.045(0.0043)	5.07E-25	TG	0.026(0.0044)	2.59E-09
cg19589396	HbA1c	0.045(0.0043)	4.09E-25	TG	0.026(0.0044)	2.39E-09
cg19695507	HbA1c	0.045(0.0043)	2.14E-25	TG	0.026(0.0044)	3.76E-09
cg19750657	HbA1c	0.042(0.0043)	1.89E-22	TG	0.025(0.0044)	3.64E-08
cg19881557	HbA1c	0.045(0.0043)	3.82E-25	TG	0.027(0.0044)	1.08E-09
cg19998073	HbA1c	0.045(0.0043)	5.35E-25	TG	0.026(0.0044)	7.89E-09
cg21108085	HbA1c	0.045(0.0043)	3.96E-26	TG	0.026(0.0044)	2.98E-09

	HbA1c					
cg21429551	HbA1c	0.045(0.0042)	1.18E-25	TG	0.027(0.0044)	1.54E-09
cg21486834	HbA1c	0.045(0.0043)	2.98E-25	TG	0.027(0.0044)	7.08E-10
cg22012981	HbA1c	0.045(0.0043)	2.39E-25	TG	0.026(0.0044)	5.28E-09
cg22103219	HbA1c	0.045(0.0043)	2.21E-25	TG	0.027(0.0044)	8.43E-10
cg22488164	HbA1c	0.044(0.0043)	4.51E-24	TG	0.027(0.0044)	7.38E-10
cg22534374	HbA1c	0.044(0.0043)	1.16E-24	TG	0.026(0.0044)	2.47E-09
cg22590032	HbA1c	0.043(0.0043)	6.16E-24	TG	0.026(0.0044)	5.97E-09
cg22695339	HbA1c	0.045(0.0043)	2.80E-25	TG	0.027(0.0044)	6.43E-10
cg22700686	HbA1c	0.045(0.0043)	4.86E-25	TG	0.027(0.0044)	1.62E-09
cg23032421	HbA1c	0.045(0.0042)	1.11E-25	TG	0.027(0.0044)	1.36E-09
cg23232188	HbA1c	0.044(0.0043)	1.04E-24	TG	0.027(0.0044)	1.26E-09
cg24174557	HbA1c	0.044(0.0043)	2.86E-24	TG	0.025(0.0044)	8.94E-09
cg24403644	HbA1c	0.045(0.0043)	5.87E-26	TG	0.026(0.0044)	2.10E-09
cg24469729	HbA1c	0.044(0.0043)	1.95E-24	TG	0.027(0.0044)	1.71E-09
cg24531955	HbA1c	0.044(0.0043)	1.02E-24	TG	0.027(0.0044)	6.83E-10
cg24679890	HbA1c	0.046(0.0043)	3.03E-26	TG	0.027(0.0044)	7.88E-10
cg25001190	HbA1c	0.044(0.0043)	1.71E-24	TG	0.026(0.0044)	4.63E-09
cg25096107	HbA1c	0.045(0.0043)	4.27E-25	TG	0.027(0.0044)	6.51E-10
cg25197194	HbA1c	0.045(0.0043)	1.86E-25	TG	0.027(0.0044)	1.82E-09
cg25217710	HbA1c	0.043(0.0043)	1.57E-23	TG	0.026(0.0044)	7.70E-09
cg25435714	HbA1c	0.044(0.0043)	8.73E-25	TG	0.026(0.0044)	3.23E-09
cg25570328	HbA1c	0.045(0.0043)	1.27E-25	TG	0.025(0.0044)	1.59E-08
cg25649826	HbA1c	0.045(0.0043)	2.18E-25	TG	0.026(0.0044)	3.39E-09
cg26033520	HbA1c	0.045(0.0043)	2.28E-25	TG	0.026(0.0044)	6.30E-09
cg26253134	HbA1c	0.045(0.0043)	2.95E-25	TG	0.026(0.0044)	5.22E-09
cg26357885	HbA1c	0.045(0.0043)	3.53E-26	TG	0.027(0.0044)	7.43E-10
cg26361535	HbA1c	0.044(0.0043)	1.29E-24	TG	0.026(0.0044)	3.33E-09

	HbA1c					
cg26403843	HbA1c	0.044(0.0043)	1.44E-24	TG	0.025(0.0044)	1.88E-08
cg26542660	HbA1c	0.045(0.0043)	1.65E-25	TG	0.027(0.0044)	1.75E-09
cg26663590	HbA1c	0.044(0.0043)	4.03E-24	TG	0.026(0.0044)	4.58E-09
cg26687842	HbA1c	0.045(0.0043)	3.24E-25	TG	0.026(0.0044)	2.56E-09
cg26804423	HbA1c	0.044(0.0043)	3.79E-24	TG	0.027(0.0044)	1.05E-09
cg26836479	HbA1c	0.045(0.0042)	8.35E-26	TG	0.026(0.0044)	2.95E-09
cg26878209	HbA1c	0.045(0.0043)	2.85E-25	TG	0.026(0.0044)	6.18E-09
cg26894079	HbA1c	0.045(0.0043)	6.24E-26	TG	0.027(0.0044)	7.39E-10
cg26952928	HbA1c	0.045(0.0043)	4.50E-25	TG	0.027(0.0044)	2.03E-09
cg27050612	HbA1c	0.046(0.0043)	2.53E-26	TG	0.026(0.0044)	3.87E-09
cg27087650	HbA1c	0.045(0.0043)	2.50E-25	TG	0.026(0.0044)	2.10E-09
cg27115863	HbA1c	0.045(0.0043)	2.95E-25	TG	0.028(0.0044)	4.68E-10
cg27117792	HbA1c	0.045(0.0043)	8.11E-26	TG	0.027(0.0044)	1.53E-09
cg27184903	HbA1c	0.044(0.0043)	1.97E-24	TG	0.027(0.0044)	6.75E-10
cg27269962	HbA1c	0.045(0.0043)	1.82E-25	TG	0.026(0.0044)	2.90E-09
cg27547344	HbA1c	0.044(0.0043)	8.36E-25	TG	0.027(0.0044)	6.46E-10
cg27614723	HbA1c	0.045(0.0043)	6.38E-26	TG	0.026(0.0044)	2.95E-09
ch.2.304154 74F	HbA1c	0.045(0.0043)	1.01E-25	TG	0.026(0.0044)	2.43E-09
cg00094412	HDL	-0.048(0.0041)	4.26E-30	WHR	0.081(0.0039)	6.84E-91
cg00108715	HDL	-0.047(0.0041)	6.02E-30	WHR	0.081(0.0039)	1.29E-91
cg00138407	HDL	-0.047(0.0041)	6.79E-30	WHR	0.080(0.0039)	3.85E-89
cg00144180	HDL	-0.048(0.0041)	3.83E-30	WHR	0.081(0.0039)	4.04E-91
cg00238353	HDL	-0.047(0.0041)	2.14E-29	WHR	0.082(0.0039)	3.85E-92
cg00244001	HDL	-0.048(0.0041)	1.72E-30	WHR	0.081(0.0039)	2.11E-91
cg00431050	HDL	-0.049(0.0041)	9.33E-32	WHR	0.082(0.0039)	3.61E-91
cg00574958	HDL	-0.048(0.0042)	1.82E-29	WHR	0.080(0.0039)	2.88E-87
cg00634542	HDL	-0.048(0.0041)	1.17E-30	WHR	0.081(0.0039)	3.90E-90
cg00673344	HDL	-0.048(0.0041)	1.30E-30	WHR	0.081(0.0038)	8.35E-92
cg00711896	HDL	-0.048(0.0041)	4.66E-31	WHR	0.081(0.0039)	9.60E-91
cg00863378	HDL	-0.047(0.0041)	1.17E-29	WHR	0.082(0.0039)	1.73E-91
cg00973118	HDL	-0.048(0.0041)	3.76E-30	WHR	0.081(0.0039)	4.41E-91
cg01101459	HDL	-0.046(0.0041)	1.67E-28	WHR	0.080(0.0039)	2.09E-88
cg01243823	HDL	-0.048(0.0041)	9.34E-31	WHR	0.081(0.0039)	1.74E-91

cg01511901	HDL	-0.049(0.0041)	1.10E-31	WHR	0.082(0.0039)	1.40E-91
cg01798813	HDL	-0.047(0.0041)	5.08E-29	WHR	0.080(0.0039)	7.06E-89
cg02119938	HDL	-0.049(0.0041)	1.80E-31	WHR	0.081(0.0039)	3.28E-91
cg02286155	HDL	-0.047(0.0041)	3.23E-29	WHR	0.081(0.0039)	6.21E-90
cg02560388	HDL	-0.046(0.0041)	1.78E-28	WHR	0.081(0.0039)	2.87E-90
cg02650017	HDL	-0.044(0.0041)	2.05E-26	WHR	0.080(0.0039)	1.83E-87
cg02711608	HDL	-0.048(0.0041)	1.24E-30	WHR	0.081(0.0038)	1.99E-91
cg02716826	HDL	-0.046(0.0041)	1.18E-28	WHR	0.081(0.0039)	2.04E-90
cg03050965	HDL	-0.048(0.0041)	1.52E-30	WHR	0.081(0.0039)	4.96E-91
cg03159676	HDL	-0.047(0.0041)	8.92E-29	WHR	0.081(0.0039)	2.90E-89
cg03318904	HDL	-0.048(0.0041)	1.41E-30	WHR	0.079(0.0039)	5.39E-87
cg03327570	HDL	-0.047(0.0041)	3.68E-29	WHR	0.082(0.0039)	1.04E-91
cg03433986	HDL	-0.048(0.0041)	5.46E-31	WHR	0.082(0.0039)	2.00E-92
cg03523676	HDL	-0.047(0.0041)	4.97E-29	WHR	0.081(0.0039)	4.72E-90
cg03725309	HDL	-0.049(0.0041)	2.42E-31	WHR	0.080(0.0038)	1.27E-89
cg03885055	HDL	-0.048(0.0041)	1.51E-30	WHR	0.081(0.0039)	2.56E-91
cg03940776	HDL	-0.047(0.0041)	8.99E-30	WHR	0.081(0.0039)	6.03E-91
cg03957124	HDL	-0.048(0.0041)	1.95E-30	WHR	0.081(0.0039)	2.96E-89
cg04011474	HDL	-0.048(0.0041)	3.20E-30	WHR	0.081(0.0039)	2.50E-90
cg04126866	HDL	-0.048(0.0041)	1.08E-30	WHR	0.081(0.0039)	6.00E-91
cg04232128	HDL	-0.048(0.0041)	3.66E-30	WHR	0.081(0.0039)	7.66E-90
cg04524040	HDL	-0.048(0.0041)	1.81E-30	WHR	0.082(0.0038)	1.24E-92
cg04577162	HDL	-0.046(0.0041)	1.62E-28	WHR	0.081(0.0039)	1.25E-89
cg05063895	HDL	-0.048(0.0041)	6.60E-31	WHR	0.082(0.0039)	2.50E-92
cg05095590	HDL	-0.047(0.0041)	4.53E-30	WHR	0.082(0.0039)	1.16E-92
cg05648472	HDL	-0.047(0.0041)	1.30E-28	WHR	0.081(0.0039)	4.87E-89
cg05720226	HDL	-0.046(0.0041)	2.66E-28	WHR	0.081(0.0039)	5.92E-91
cg05845030	HDL	-0.047(0.0041)	2.16E-29	WHR	0.081(0.0039)	1.69E-90
cg06012428	HDL	-0.047(0.0041)	8.86E-30	WHR	0.082(0.0039)	1.05E-91
cg06164260	HDL	-0.047(0.0041)	7.34E-29	WHR	0.082(0.0039)	2.33E-91
cg06192883	HDL	-0.047(0.0042)	1.95E-28	WHR	0.079(0.0039)	1.93E-86
cg06500161	HDL	-0.039(0.0041)	4.85E-21	WHR	0.079(0.0039)	6.04E-85
cg06559575	HDL	-0.047(0.0041)	1.89E-29	WHR	0.082(0.0039)	1.28E-91
cg06603309	HDL	-0.045(0.0041)	3.00E-27	WHR	0.081(0.0039)	1.61E-90
cg06690548	HDL	-0.048(0.0041)	1.97E-30	WHR	0.081(0.0039)	9.35E-91
cg06898549	HDL	-0.047(0.0041)	3.11E-29	WHR	0.081(0.0039)	1.05E-90
cg06946797	HDL	-0.047(0.0041)	2.09E-29	WHR	0.081(0.0039)	7.16E-89
cg07021906	HDL	-0.048(0.0041)	3.27E-30	WHR	0.081(0.0039)	6.29E-90
cg07037944	HDL	-0.048(0.0041)	1.47E-30	WHR	0.081(0.0039)	6.61E-90
cg07136133	HDL	-0.048(0.0041)	5.57E-30	WHR	0.081(0.0039)	7.49E-91
cg07202479	HDL	-0.047(0.0041)	6.70E-30	WHR	0.081(0.0039)	9.06E-91
cg07471614	HDL	-0.048(0.0041)	3.09E-30	WHR	0.082(0.0039)	2.46E-91
cg07504977	HDL	-0.047(0.0041)	3.64E-29	WHR	0.081(0.0039)	1.52E-90
cg07682160	HDL	-0.045(0.0041)	2.48E-27	WHR	0.081(0.0039)	7.89E-90
cg07728579	HDL	-0.046(0.0041)	1.17E-28	WHR	0.081(0.0039)	3.43E-90
cg07769588	HDL	-0.046(0.0041)	1.05E-27	WHR	0.081(0.0039)	1.05E-89
cg08305942	HDL	-0.048(0.0041)	1.59E-30	WHR	0.082(0.0039)	2.62E-92
cg08309687	HDL	-0.047(0.0041)	2.52E-29	WHR	0.081(0.0039)	1.77E-90

cg08443038	HDL	-0.048(0.0041)	1.37E-30	WHR	0.082(0.0039)	5.13E-92
cg08548559	HDL	-0.048(0.0041)	1.19E-30	WHR	0.081(0.0039)	1.99E-91
cg08648047	HDL	-0.047(0.0041)	1.79E-29	WHR	0.082(0.0039)	2.07E-91
cg08726900	HDL	-0.047(0.0041)	8.08E-29	WHR	0.080(0.0039)	5.04E-89
cg08813944	HDL	-0.049(0.0041)	2.17E-31	WHR	0.081(0.0039)	6.48E-91
cg08857797	HDL	-0.047(0.0041)	2.84E-29	WHR	0.081(0.0039)	4.91E-90
cg09152259	HDL	-0.047(0.0041)	1.26E-29	WHR	0.081(0.0039)	6.85E-91
cg09222732	HDL	-0.047(0.0041)	7.09E-30	WHR	0.081(0.0039)	2.77E-90
cg09315878	HDL	-0.048(0.0041)	1.77E-30	WHR	0.080(0.0039)	2.06E-89
cg09349128	HDL	-0.045(0.0042)	3.71E-26	WHR	0.079(0.0039)	1.12E-85
cg09554443	HDL	-0.045(0.0041)	1.01E-27	WHR	0.081(0.0039)	3.16E-90
cg09613192	HDL	-0.047(0.0041)	4.37E-29	WHR	0.081(0.0039)	9.94E-91
cg09664445	HDL	-0.047(0.0042)	2.50E-29	WHR	0.081(0.0039)	1.47E-88
cg09777883	HDL	-0.047(0.0041)	8.54E-30	WHR	0.081(0.0039)	6.18E-91
cg10179300	HDL	-0.046(0.0041)	1.97E-28	WHR	0.082(0.0039)	3.50E-91
cg10438589	HDL	-0.047(0.0041)	2.29E-29	WHR	0.081(0.0039)	2.01E-91
cg10505902	HDL	-0.047(0.0041)	1.25E-29	WHR	0.081(0.0039)	2.14E-91
cg10513161	HDL	-0.048(0.0041)	1.76E-30	WHR	0.081(0.0039)	3.77E-90
cg10549088	HDL	-0.047(0.0041)	1.25E-29	WHR	0.081(0.0039)	5.11E-91
cg10717869	HDL	-0.047(0.0041)	9.61E-30	WHR	0.082(0.0039)	5.73E-92
cg10734665	HDL	-0.047(0.0041)	5.70E-29	WHR	0.081(0.0039)	2.09E-90
cg10814005	HDL	-0.047(0.0041)	1.20E-29	WHR	0.081(0.0039)	2.61E-91
cg10919522	HDL	-0.047(0.0041)	1.55E-29	WHR	0.081(0.0039)	2.16E-90
cg10922280	HDL	-0.046(0.0041)	4.88E-28	WHR	0.081(0.0039)	2.65E-89
cg10927968	HDL	-0.047(0.0041)	3.56E-29	WHR	0.080(0.0039)	4.11E-88
cg10975897	HDL	-0.047(0.0041)	2.30E-29	WHR	0.081(0.0039)	1.49E-90
cg11024682	HDL	-0.045(0.0042)	6.14E-27	WHR	0.079(0.0039)	3.53E-86
cg11080651	HDL	-0.048(0.0041)	9.83E-31	WHR	0.081(0.0039)	1.54E-91
cg11183227	HDL	-0.048(0.0041)	9.05E-31	WHR	0.081(0.0039)	7.91E-90
cg11202345	HDL	-0.047(0.0041)	1.01E-29	WHR	0.081(0.0039)	2.17E-89
cg11376147	HDL	-0.049(0.0041)	1.97E-31	WHR	0.081(0.0039)	5.89E-90
cg11614585	HDL	-0.045(0.0041)	3.03E-27	WHR	0.081(0.0039)	9.06E-90
cg11650298	HDL	-0.047(0.0041)	2.67E-29	WHR	0.081(0.0039)	1.87E-90
cg11832534	HDL	-0.047(0.0041)	1.41E-29	WHR	0.080(0.0039)	2.15E-89
cg11927233	HDL	-0.047(0.0041)	5.74E-30	WHR	0.082(0.0039)	3.86E-92
cg11969813	HDL	-0.047(0.0041)	5.05E-29	WHR	0.080(0.0039)	2.71E-88
cg12484113	HDL	-0.046(0.0041)	9.66E-28	WHR	0.081(0.0039)	5.11E-90
cg12593793	HDL	-0.048(0.0042)	1.16E-30	WHR	0.080(0.0039)	5.36E-88
cg12992827	HDL	-0.045(0.0041)	3.06E-27	WHR	0.080(0.0039)	3.32E-88
cg13097800	HDL	-0.048(0.0041)	9.35E-31	WHR	0.082(0.0039)	2.09E-92
cg13123009	HDL	-0.047(0.0041)	1.16E-29	WHR	0.081(0.0039)	1.84E-90
cg13274938	HDL	-0.046(0.0041)	1.21E-28	WHR	0.081(0.0039)	5.07E-90
cg13591783	HDL	-0.048(0.0041)	1.16E-30	WHR	0.082(0.0039)	3.05E-92
cg13781414	HDL	-0.046(0.0041)	4.24E-28	WHR	0.082(0.0039)	1.88E-92
cg13922488	HDL	-0.047(0.0041)	5.79E-29	WHR	0.081(0.0039)	6.95E-91
cg14020176	HDL	-0.048(0.0041)	3.22E-30	WHR	0.081(0.0039)	9.31E-91
cg14264316	HDL	-0.048(0.0041)	2.37E-30	WHR	0.081(0.0039)	2.40E-90
cg14476101	HDL	-0.047(0.0041)	1.06E-29	WHR	0.080(0.0039)	5.60E-89

cg15323828	HDL	-0.048(0.0041)	1.25E-30	WHR	0.082(0.0039)	5.63E-92
cg15357118	HDL	-0.047(0.0041)	6.20E-29	WHR	0.082(0.0039)	3.48E-92
cg15681239	HDL	-0.048(0.0041)	5.36E-31	WHR	0.081(0.0039)	8.24E-91
cg15721584	HDL	-0.045(0.0041)	4.61E-27	WHR	0.081(0.0039)	2.00E-89
cg16163382	HDL	-0.047(0.0041)	8.23E-29	WHR	0.081(0.0039)	6.22E-90
cg16578636	HDL	-0.048(0.0041)	2.43E-30	WHR	0.081(0.0038)	7.07E-92
cg16594806	HDL	-0.047(0.0042)	2.48E-29	WHR	0.081(0.0039)	2.37E-89
cg16611584	HDL	-0.046(0.0041)	3.04E-28	WHR	0.081(0.0039)	1.66E-90
cg16815882	HDL	-0.047(0.0041)	1.32E-29	WHR	0.081(0.0039)	9.25E-90
cg16846518	HDL	-0.049(0.0041)	1.84E-31	WHR	0.082(0.0039)	3.44E-92
cg17178175	HDL	-0.047(0.0041)	1.28E-29	WHR	0.081(0.0039)	7.16E-91
cg17260706	HDL	-0.046(0.0041)	1.92E-28	WHR	0.081(0.0039)	6.43E-90
cg17501210	HDL	-0.045(0.0041)	3.34E-27	WHR	0.080(0.0039)	8.28E-88
cg17901584	HDL	-0.043(0.0041)	4.93E-25	WHR	0.081(0.0039)	6.83E-89
cg17971578	HDL	-0.048(0.0041)	6.81E-31	WHR	0.081(0.0038)	3.95E-92
cg18098839	HDL	-0.047(0.0041)	1.10E-29	WHR	0.081(0.0039)	4.22E-90
cg18120259	HDL	-0.048(0.0041)	2.79E-30	WHR	0.081(0.0038)	2.98E-91
cg18181703	HDL	-0.046(0.0041)	1.03E-27	WHR	0.080(0.0039)	3.04E-88
cg18217136	HDL	-0.049(0.0041)	3.25E-31	WHR	0.081(0.0039)	2.51E-90
cg18219562	HDL	-0.047(0.0041)	1.10E-29	WHR	0.081(0.0039)	4.11E-91
cg18513344	HDL	-0.046(0.0041)	1.13E-27	WHR	0.081(0.0039)	2.76E-89
cg18608055	HDL	-0.045(0.0041)	8.52E-27	WHR	0.081(0.0039)	3.56E-89
cg19217955	HDL	-0.048(0.0041)	1.75E-30	WHR	0.081(0.0039)	2.16E-90
cg19373099	HDL	-0.046(0.0041)	4.58E-28	WHR	0.081(0.0039)	2.86E-90
cg19566658	HDL	-0.046(0.0041)	2.43E-28	WHR	0.081(0.0039)	4.10E-91
cg19589396	HDL	-0.048(0.0041)	2.16E-30	WHR	0.081(0.0039)	3.13E-90
cg19695507	HDL	-0.046(0.0041)	4.30E-28	WHR	0.081(0.0039)	8.70E-90
cg19750657	HDL	-0.046(0.0042)	1.11E-27	WHR	0.080(0.0039)	1.56E-86
cg19881557	HDL	-0.046(0.0041)	2.74E-28	WHR	0.080(0.0039)	7.83E-88
cg19998073	HDL	-0.047(0.0041)	3.81E-29	WHR	0.082(0.0039)	2.89E-91
cg21108085	HDL	-0.047(0.0041)	5.06E-30	WHR	0.081(0.0039)	2.60E-91
cg21429551	HDL	-0.048(0.0041)	7.95E-31	WHR	0.081(0.0038)	1.92E-91
cg21486834	HDL	-0.047(0.0041)	4.63E-29	WHR	0.081(0.0039)	1.72E-89
cg22012981	HDL	-0.047(0.0041)	1.94E-29	WHR	0.082(0.0039)	4.08E-92
cg22103219	HDL	-0.048(0.0041)	1.37E-30	WHR	0.081(0.0038)	8.48E-91
cg22488164	HDL	-0.045(0.0041)	5.78E-27	WHR	0.080(0.0039)	4.28E-88
cg22534374	HDL	-0.046(0.0041)	3.77E-28	WHR	0.081(0.0039)	4.06E-90
cg22590032	HDL	-0.045(0.0041)	1.55E-27	WHR	0.081(0.0039)	5.14E-90
cg22695339	HDL	-0.047(0.0041)	5.36E-29	WHR	0.081(0.0039)	3.86E-90
cg22700686	HDL	-0.047(0.0041)	1.93E-29	WHR	0.081(0.0039)	5.44E-90
cg23032421	HDL	-0.048(0.0041)	1.40E-30	WHR	0.082(0.0038)	1.71E-92
cg23232188	HDL	-0.048(0.0041)	8.23E-30	WHR	0.080(0.0039)	1.13E-88
cg24174557	HDL	-0.047(0.0041)	3.62E-29	WHR	0.080(0.0039)	5.84E-89
cg24403644	HDL	-0.048(0.0041)	5.92E-31	WHR	0.081(0.0039)	4.13E-91
cg24469729	HDL	-0.047(0.0041)	2.61E-29	WHR	0.082(0.0039)	3.65E-91
cg24531955	HDL	-0.048(0.0041)	3.06E-30	WHR	0.081(0.0039)	1.93E-90
cg24679890	HDL	-0.048(0.0041)	1.08E-30	WHR	0.081(0.0039)	2.71E-91
cg25001190	HDL	-0.046(0.0041)	6.69E-29	WHR	0.081(0.0039)	8.51E-91

cg25096107	HDL	-0.045(0.0041)	2.82E-27	WHR	0.081(0.0039)	4.87E-90
cg25197194	HDL	-0.047(0.0041)	1.58E-29	WHR	0.081(0.0039)	7.43E-92
cg25217710	HDL	-0.047(0.0041)	1.36E-29	WHR	0.080(0.0039)	1.90E-88
cg25435714	HDL	-0.046(0.0041)	2.38E-28	WHR	0.081(0.0039)	4.42E-90
cg25570328	HDL	-0.047(0.0041)	7.11E-29	WHR	0.081(0.0039)	2.41E-91
cg25649826	HDL	-0.046(0.0041)	9.61E-29	WHR	0.081(0.0039)	9.77E-90
cg26033520	HDL	-0.045(0.0041)	1.56E-27	WHR	0.081(0.0039)	8.86E-90
cg26253134	HDL	-0.047(0.0041)	9.19E-30	WHR	0.081(0.0039)	8.81E-91
cg26357885	HDL	-0.049(0.0041)	5.74E-32	WHR	0.082(0.0039)	2.56E-92
cg26361535	HDL	-0.046(0.0041)	2.02E-28	WHR	0.082(0.0039)	1.73E-91
cg26403843	HDL	-0.046(0.0041)	2.13E-28	WHR	0.081(0.0039)	1.21E-90
cg26542660	HDL	-0.048(0.0041)	4.28E-30	WHR	0.081(0.0038)	4.25E-90
cg26663590	HDL	-0.048(0.0041)	6.47E-30	WHR	0.081(0.0039)	1.42E-89
cg26687842	HDL	-0.048(0.0041)	1.55E-30	WHR	0.080(0.0038)	3.52E-89
cg26804423	HDL	-0.046(0.0041)	4.65E-28	WHR	0.081(0.0039)	6.82E-90
cg26836479	HDL	-0.048(0.0041)	2.00E-30	WHR	0.081(0.0038)	6.62E-92
cg26878209	HDL	-0.046(0.0041)	4.52E-28	WHR	0.081(0.0039)	1.35E-89
cg26894079	HDL	-0.048(0.0041)	3.04E-30	WHR	0.082(0.0039)	1.65E-91
cg26952928	HDL	-0.047(0.0041)	5.11E-29	WHR	0.081(0.0039)	5.75E-90
cg27050612	HDL	-0.048(0.0041)	1.35E-30	WHR	0.081(0.0039)	1.01E-90
cg27087650	HDL	-0.047(0.0041)	6.95E-30	WHR	0.081(0.0039)	1.12E-90
cg27115863	HDL	-0.049(0.0041)	2.58E-31	WHR	0.082(0.0039)	4.60E-92
cg27117792	HDL	-0.048(0.0041)	4.00E-30	WHR	0.082(0.0039)	2.81E-92
cg27184903	HDL	-0.048(0.0041)	3.40E-30	WHR	0.081(0.0039)	1.22E-90
cg27269962	HDL	-0.047(0.0041)	7.00E-30	WHR	0.081(0.0039)	3.98E-91
cg27547344	HDL	-0.047(0.0041)	4.52E-29	WHR	0.081(0.0039)	2.09E-91
cg27614723	HDL	-0.047(0.0041)	1.17E-29	WHR	0.082(0.0039)	4.69E-92
ch.2.304154 74F	HDL	-0.047(0.0041)	8.31E-30	WHR	0.081(0.0038)	4.84E-91

Effect (SE): Effect (SE) of the CpG effect in the meta-analyses, P-value: p-value of the CpG effect in the meta-analyses**,** Chol: total cholesterol, CRP: C-reactive protein,Glc: glucose, HDL: high-density lipoprotein cholesterol, Ins: insulin, LDL: low-density lipoprotein cholesterol, SBP**:** systolic blood pressure, TG**:** triglycerides**,** WHR: wasit-hip ratio

Appendix Table 4: T2D and obesity SNPs analysed for HSM effects

T2D			Obesity		
Chromosome	**Position**	**rs ID**	**Chromosome**	**Position**	**rs ID**
4	6292915	rs10010131	1	72812440	rs2815752
19	19407718	rs10401969	1	74977870	rs1040070
6	20688121	rs10440833	1	74991644	rs1514175
9	22134094	rs10811661	1	78048331	rs17381664
11	92708710	rs10830963	1	96944797	rs1555543
12	27965150	rs10842994	1	112264540	rs2076592
1	120517959	rs10923931	1	119503843	rs984222
9	22133284	rs10965250	1	172346548	rs1011731
12	4374373	rs11063069	1	177855517	rs516636
10	94462882	rs1111875	1	177873210	rs574367
15	80432222	rs11634397	1	177889480	rs543874
16	53845487	rs11642841	1	177913519	rs10913469
3	123065778	rs11708067	1	219644224	rs2605100
18	58049192	rs11873305	1	219657163	rs11118316
2	43618819	rs11899863	2	622827	rs2867125
10	80942631	rs12571751	2	638144	rs4854344
10	12328010	rs12779790	2	12964497	rs10198628
18	57884750	rs12970134	2	25131316	rs6545814
3	12289800	rs13081389	2	25150116	rs6752378
8	118184783	rs13266634	2	25158008	rs713586
9	81952128	rs13292136	2	59302877	rs887912
2	165528876	rs13389219	2	142959931	rs2890652
13	80717156	rs1359790	2	165513091	rs10195252
11	92673828	rs1387153	2	227105921	rs2943650
3	185529080	rs1470579	3	12488882	rs4684854
12	66174894	rs1531343	3	64705365	rs6795735
6	39284050	rs1535500	3	85884150	rs13078807
11	72433098	rs1552224	3	185834499	rs9816226
11	2847069	rs163184	4	45182527	rs10938397
3	186666461	rs16861329	4	103188709	rs13107325
7	14898282	rs17168486	4	130731284	rs4864201
9	8879118	rs17584499	5	55859952	rs11743303
4	6303022	rs1801214	5	75015242	rs2112347
3	12393125	rs1801282	5	95850250	rs261967
10	70931474	rs1802295	5	124332103	rs4836133
15	90374257	rs2028299	5	173362458	rs6861681
7	15064309	rs2191349	6	6743149	rs1294421
11	2691500	rs231361	6	20685486	rs9356744
11	2691471	rs231362	6	20694884	rs2206734
11	1696849	rs2334499	6	34302869	rs206936
2	60584819	rs243021	6	43758873	rs6905288

2	60568745	rs243088
9	84308948	rs2796441
2	227093745	rs2943641
1	214159256	rs340874
19	33893008	rs3786897
8	118185025	rs3802177
17	2216258	rs391300
2	165501849	rs3923113
17	36098040	rs4430796
5	76424949	rs4457053
5	55806751	rs459193
3	64711904	rs4607103
7	44235668	rs4607517
12	71634794	rs4760790
20	42989267	rs4812829
10	94465559	rs5015480
8	41519248	rs516946
11	17408630	rs5215
X	152899922	rs5945326
20	42946966	rs6017317
7	127164958	rs6467136
3	64705365	rs6795735
7	15052860	rs6960043
9	4287466	rs7041847
15	62391608	rs7163757
15	77832762	rs7177055
15	77747190	rs7178572
16	75247245	rs7202877
2	227020653	rs7578326
2	43732823	rs7578597
2	161171454	rs7593730
3	23336450	rs7612463
6	20661250	rs7754840
2	27741237	rs780094
10	114758349	rs7903146
12	121460686	rs7957197
15	91521337	rs8042680
19	46158513	rs8108269
3	64048297	rs831571
7	28196222	rs849134
8	95960511	rs896854
9	22051670	rs944801
6	38106844	rs9470794
7	130466854	rs972283
6	50786571	rs2272903
6	50803050	rs987237
6	50836279	rs734597
6	127452639	rs9491696
7	25871109	rs1055144
8	9745798	rs17150703
8	76615663	rs4735692
9	28414339	rs10968576
9	36969205	rs16933812
9	72998332	rs11142387
9	87285992	rs1211166
10	31990623	rs7081678
10	78646536	rs2116830
11	8604593	rs4929949
11	27563382	rs988712
11	27679916	rs6265
11	27725986	rs10767664
11	27728539	rs2030323
11	47650993	rs3817334
12	26453283	rs718314
12	50247468	rs7138803
12	50263148	rs7132908
12	54342684	rs1443512
13	28020180	rs4771122
13	54064981	rs9568856
13	80959207	rs534870
13	97017548	rs7989336
14	78786077	rs11624704
14	79936964	rs10150332
14	79945162	rs10146997
15	68016343	rs997295
15	68080886	rs4776970
16	4013467	rs2531995
16	19933600	rs12444979
16	20257867	rs12597579
16	28855727	rs4788099
16	28885659	rs7359397
16	53800954	rs1421085
16	53803574	rs1558902
16	53813367	rs17817449
16	53816275	rs8050136
16	53818460	rs3751812
16	53819169	rs9936385
16	53820527	rs9939609
16	53825488	rs9941349
16	53842908	rs12149832

	17	46669430	rs9299
	17	78591111	rs7503807
	18	57811982	rs17700144
	18	57828637	rs2331841
	18	57829135	rs6567160
	18	57838401	rs663129
	18	57839769	rs571312
	18	57851097	rs17782313
	18	57884750	rs12970134
	19	19329924	rs2228603
	19	34309532	rs29941
	19	45395619	rs2075650
	19	46172278	rs11671664
	19	46202172	rs2287019
	19	47569003	rs3810291
	20	51092996	rs13041126
	22	29451671	rs4823006
	22	42236337	rs5996074
	22	44324727	rs738409

References

1. Aguiree F, et al.: IDF Diabetes Atlas. 2013.
2. Whiting DR, et al.: IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res Clin Pract, 2011. 94(3): p. 311-321.
3. Roglic G and Unwin N: Mortality attributable to diabetes: Estimates for the year 2010. Diabetes Res Clin Pract, 2010. 87(1): p. 15-19.
4. Wang YC, et al.: Health and economic burden of the projected obesity trends in the USA and the UK. The Lancet, 2011. 378(9793): p. 815-825.
5. Ng M, et al.: Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet, 2014.
6. IDF: IDF Worldwide Definition of the Metabolic Syndrome. 2009.
7. Szablewski L: Role of immune system in type 1 diabetes mellitus pathogenesis. Int Immunopharmacol, 2014. 22(1): p. 182-191.
8. Rubarth LB: Infants of diabetic mothers. Neonatal Netw, 2013. 32(6): p. 416-8.
9. Thanabalasingham G and Owen KR: Diagnosis and management of maturity onset diabetes of the young (MODY). BMJ, 2011. 343: p. d6044.
10. Consultation WHO, *Definition, diagnosis and classification of diabetes mellitus and its complications*. Vol. 1. 1999: Part.
11. American Diabetes Association: Diagnosis and classification of diabetes mellitus. Diabetes Care, 2013. 36 Suppl 1: p. S67-74.
12. Colagiuri S: Glycated haemoglobin (HbA1c) for the diagnosis of diabetes mellitus--practical implications. Diabetes Res Clin Pract, 2011. 93(3): p. 312-3.
13. Hare MJ, et al.: Glucose-independent ethnic differences in HbA1c in people without known diabetes. Diabetes Care, 2013. 36(6): p. 1534-40.
14. Centers for Disease C and Prevention: Awareness of prediabetes--United States, 2005-2010. MMWR Morb Mortal Wkly Rep, 2013. 62(11): p. 209-12.
15. Abdullah N, et al.: The Architecture of Risk for Type 2 Diabetes: Understanding Asia in the Context of Global Findings. International Journal of Endocrinology, 2014. 2014: p. 21.
16. Morrish NJ, et al.: Mortality and causes of death in the WHO Multinational Study of Vascular Disease in Diabetes. Diabetologia, 2001. 44 Suppl 2: p. S14-21.
17. Young BA, Maynard C, and Boyko EJ: Racial differences in diabetic nephropathy, cardiovascular disease, and mortality in a national population of veterans. Diabetes care, 2003. 26(8): p. 2392-2399.
18. Pascolini D and Mariotti SP: Global estimates of visual impairment: 2010. British Journal of Ophthalmology, 2011: p. bjophthalmol-2011-300539.
19. Holman RR, et al.: 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med, 2008. 359(15): p. 1577-89.
20. Turner RC, et al.: Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS: 23). BMJ, 1998. 316(7134): p. 823-8.
21. Singh N, Armstrong DG, and Lipsky BA: Preventing foot ulcers in patients with diabetes. JAMA, 2005. 293(2): p. 217-28.
22. Schaper N, Apelqvist J, and Bakker K: The International Consensus and practical guidelines on the management and prevention of the diabetic foot. Current Diabetes Reports, 2003. 3(6): p. 475-479.
23. Haslam DW and James WP: Obesity. Lancet, 2005. 366(9492): p. 1197-209.

24. Bloomgarden ZT: Prevention of Obesity and Diabetes. Diabetes Care, 2003. 26(11): p. 3172-3178.
25. Sweeting H: Measurement and Definitions of Obesity In Childhood and Adolescence: A field guide for the uninitiated. Nutrition Journal, 2007. 6(1): p. 32.
26. Mei Z, et al.: Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. Am J Clin Nutr, 2002. 75(6): p. 978-85.
27. Cook Z, et al.: Use of BMI in the assessment of undernutrition in older subjects: reflecting on practice. Proc Nutr Soc, 2005. 64(3): p. 313-7.
28. Ford ES, Mokdad AH, and Giles WH: Trends in waist circumference among U.S. adults. Obes Res, 2003. 11(10): p. 1223-31.
29. Manolopoulos KN, Karpe F, and Frayn KN: Gluteofemoral body fat as a determinant of metabolic health. Int J Obes (Lond), 2010. 34(6): p. 949-59.
30. Rothman KJ: BMI-related errors in the measurement of obesity. Int J Obes (Lond), 2008. 32 Suppl 3: p. S56-9.
31. Wosje KS, Knipstein BL, and Kalkwarf HJ: Measurement error of DXA: interpretation of fat and lean mass changes in obese and non-obese children. J Clin Densitom, 2006. 9(3): p. 335-40.
32. Masters RK, Powers DA, and Link BG: Obesity and US Mortality Risk Over the Adult Life Course. American Journal of Epidemiology, 2013. 177(5): p. 431-442.
33. Alberti KG, et al.: Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation, 2009. 120(16): p. 1640-5.
34. Goldberg RB and Mather K: Targeting the consequences of the metabolic syndrome in the Diabetes Prevention Program. Arterioscler Thromb Vasc Biol, 2012. 32(9): p. 2077-90.
35. Meigs JB, Cupples LA, and Wilson PW: Parental transmission of type 2 diabetes: the Framingham Offspring Study. Diabetes, 2000. 49(12): p. 2201-2207.
36. Poulsen P, et al.: Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia, 1999. 42(2): p. 139-45.
37. Stunkard AJ, Foch TT, and Hrubec Z: A twin study of human obesity. JAMA, 1986. 256(1): p. 51-4.
38. Kobberling J and Tillil H, *Empirical risk figures for first degree relatives of non-insulin dependent diabetics*, in *The genetics of diabetes mellitus*, J. Kobberling and R. Tattersall, Editors. 1982, Academic Press: London. p. 201-209.
39. Visscher PM, Hill WG, and Wray NR: Heritability in the genomics era – concepts and misconceptions. Nat Rev Genet, 2008. 9(4): p. 255-266.
40. Poehlman ET, et al.: Heredity and changes in body composition and adipose tissue metabolism after short-term exercise-training. Eur J Appl Physiol Occup Physiol, 1987. 56(4): p. 398-402.
41. Bouchard C, et al.: The Response to Long-Term Overfeeding in Identical Twins. New England Journal of Medicine, 1990. 322(21): p. 1477-1482.
42. Stunkard AJ, et al.: The body-mass index of twins who have been reared apart. N Engl J Med, 1990. 322(21): p. 1483-7.
43. Chakravarthy MV and Booth FW: Eating, exercise, and "thrifty" genotypes: connecting the dots toward an evolutionary understanding of modern chronic diseases. J Appl Physiol, 2004. 96(1): p. 3-10.

44. Johnson PC, et al.: Intergenerational change and familial aggregation of body mass index. Eur J Epidemiol, 2012. 27(1): p. 53-61.
45. Sicree R and Shaw J: Type 2 diabetes: An epidemic or not, and why it is happening. Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 2007. 1(2): p. 75-81.
46. Shaw JE, Sicree RA, and Zimmet PZ: Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract, 2010. 87(1): p. 4-14.
47. Angelakis E, et al.: The relationship between gut microbiota and weight gain in humans. Future Microbiol, 2012. 7(1): p. 91-109.
48. Furrow RE, Christiansen FB, and Feldman MW: Environment-sensitive epigenetics and the heritability of complex diseases. Genetics, 2011. 189(4): p. 1377-87.
49. Dabelea D and Pettitt DJ: Intrauterine diabetic environment confers risks for type 2 diabetes mellitus and obesity in the offspring, in addition to genetic susceptibility. J Pediatr Endocrinol Metab, 2001. 14(8): p. 1085-91.
50. Pettitt DJ, et al.: Excessive Obesity in Offspring of Pima Indian Women with Diabetes during Pregnancy. New England Journal of Medicine, 1983. 308(5): p. 242-245.
51. Pettitt DJ, et al.: Congenital susceptibility to NIDDM. Role of intrauterine environment. Diabetes, 1988. 37(5): p. 622-8.
52. Hales CN and Barker DJ: Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. Diabetologia, 1992. 35(7): p. 595-601.
53. McCarthy MI: Genomics, type 2 diabetes, and obesity. N Engl J Med, 2010. 363(24): p. 2339-50.
54. Zhang Y, et al.: Positional cloning of the mouse obese gene and its human homologue. Nature, 1994. 372(6505): p. 425-32.
55. Chatzigeorgiou A, et al.: The use of animal models in the study of diabetes mellitus. In Vivo, 2009. 23(2): p. 245-58.
56. Hinney A, Vogel CI, and Hebebrand J: From monogenic to polygenic obesity: recent advances. Eur Child Adolesc Psychiatry, 2010. 19(3): p. 297-310.
57. Huszar D, et al.: Targeted disruption of the melanocortin-4 receptor results in obesity in mice. Cell, 1997. 88(1): p. 131-41.
58. Bonnefond A, Froguel P, and Vaxillaire M: The emerging genetics of type 2 diabetes. Trends in Molecular Medicine, 2010. 16(9): p. 407-416.
59. Muoio DM and Newgard CB: Mechanisms of disease: molecular and metabolic mechanisms of insulin resistance and beta-cell failure in type 2 diabetes. Nat Rev Mol Cell Biol, 2008. 9(3): p. 193-205.
60. Shield JP: Neonatal diabetes: new insights into aetiology and implications. Horm Res, 2000. 53 Suppl 1: p. 7-11.
61. Temple IK and Shield JP: 6q24 transient neonatal diabetes. Rev Endocr Metab Disord, 2010. 11(3): p. 199-204.
62. Gloyn AL, et al.: Permanent neonatal diabetes due to paternal germline mosaicism for an activating mutation of the KCNJ11 Gene encoding the Kir6.2 subunit of the beta-cell potassium adenosine triphosphate channel. J Clin Endocrinol Metab, 2004. 89(8): p. 3932-5.
63. Gloyn AL, et al.: Large-Scale Association Studies of Variants in Genes Encoding the Pancreatic β-Cell KATP Channel Subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) Confirm That the KCNJ11 E23K Variant Is Associated With Type 2 Diabetes. Diabetes, 2003. 52(2): p. 568-572.
64. Edghill EL, Flanagan SE, and Ellard S: Permanent neonatal diabetes due to activating mutations in ABCC8 and KCNJ11. Rev Endocr Metab Disord, 2010. 11(3): p. 193-8.
65. Edghill EL, et al.: Insulin mutation screening in 1,044 patients with diabetes: mutations in the INS gene are a common cause of neonatal diabetes but a

rare cause of diabetes diagnosed in childhood or adulthood. Diabetes, 2008. 57(4): p. 1034-42.
66. Edghill EL, et al.: Hepatocyte nuclear factor-1 beta mutations cause neonatal diabetes and intrauterine growth retardation: support for a critical role of HNF-1beta in human pancreatic development. Diabet Med, 2006. 23(12): p. 1301-6.
67. Butler MG: Genomic imprinting disorders in humans: a mini-review. J Assist Reprod Genet, 2009. 26(9-10): p. 477-86.
68. O'Rahilly S: Leptin: Defining its role in humans by the clinical study of genetic disorders. Nutr Rev, 2002. 60(10 I): p. S30-S34.
69. Farooqi IS, et al.: Heterozygosity for a POMC-null mutation and increased obesity risk in humans. Diabetes, 2006. 55(9): p. 2549-53.
70. Cooke D and Bloom S: The obesity pipeline: current strategies in the development of anti-obesity drugs. Nat Rev Drug Discov, 2006. 5(11): p. 919-31.
71. Montague CT, et al.: Congenital leptin deficiency is associated with severe early-onset obesity in humans. Nature, 1997. 387(6636): p. 903-8.
72. Farooqi IS, et al.: Effects of recombinant leptin therapy in a child with congenital leptin deficiency. N Engl J Med, 1999. 341(12): p. 879-84.
73. Smith EY, et al.: Transcription Is Required to Establish Maternal Imprinting at the Prader-Willi Syndrome and Angelman Syndrome Locus. PLoS Genet, 2011. 7(12): p. e1002422.
74. Weinstein LS, et al.: The role of GNAS and other imprinted genes in the development of obesity. Int J Obes (Lond), 2010. 34(1): p. 6-17.
75. Rankinen T, et al.: The human obesity gene map: the 2005 update. Obesity (Silver Spring), 2006. 14(4): p. 529-644.
76. Ansley SJ, et al.: Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. Nature, 2003. 425(6958): p. 628-33.
77. O'Rahilly S: Human genetics illuminates the paths to metabolic disease. Nature, 2009. 462(7271): p. 307-14.
78. Hattersley AT and McCarthy MI: What makes a good genetic association study? Lancet, 2005. 366(9493): p. 1315-23.
79. Altshuler D, et al.: The common PPAR[gamma] Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet, 2000. 26(1): p. 76-80.
80. Larsen LH, et al.: Prevalence of mutations and functional analyses of melanocortin 4 receptor variants identified among 750 men with juvenile-onset obesity. J Clin Endocrinol Metab, 2005. 90(1): p. 219-24.
81. Dawn Teare M and Barrett JH: Genetic linkage studies. Lancet, 2005. 366(9490): p. 1036-44.
82. Risch N and Merikangas K: The future of genetic studies of complex human diseases. Science, 1996. 273(5281): p. 1516-7.
83. Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 2007. 447(7145): p. 661-78.
84. Zeggini E, et al.: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet, 2008. 40(5): p. 638-45.
85. Dupuis J, et al.: New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet, 2010. 42(2): p. 105-16.
86. Speliotes EK, et al.: Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet, 2010. 42(11): p. 937-948.

87. Heid IM, et al.: Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet, 2010. 42(11): p. 949-960.
88. Kong A, et al.: Parental origin of sequence variants associated with complex diseases. Nature, 2009. 462(7275): p. 868-74.
89. Morris AP, et al.: Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet, 2012. 44(9): p. 981-990.
90. Voight BF, et al.: Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet, 2010. 42(7): p. 579-89.
91. Cho YS, et al.: Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet, 2012. 44(1): p. 67-72.
92. Li H, et al.: A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. Diabetes, 2013. 62(1): p. 291-8.
93. Qi L, et al.: Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Hum Mol Genet, 2010. 19(13): p. 2706-15.
94. Takeuchi F, et al.: Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. Diabetes, 2009. 58(7): p. 1690-9.
95. Yamauchi T, et al.: A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nat Genet, 2010. 42(10): p. 864-8.
96. Kooner JS, et al.: Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet, 2011. 43(10): p. 984-9.
97. Tabassum R, et al.: Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. Diabetes, 2013. 62(3): p. 977-86.
98. Parra EJ, et al.: Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County, Texas. Diabetologia, 2011. 54(8): p. 2038-46.
99. Palmer ND, et al.: A genome-wide association search for type 2 diabetes genes in African Americans. PLoS One, 2012. 7(1): p. e29202.
100. Tsai FJ, et al.: A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. PLoS Genet, 2010. 6(2): p. e1000847.
101. Grant SF, et al.: Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet, 2006. 38(3): p. 320-3.
102. Shu XO, et al.: Identification of new genetic risk variants for type 2 diabetes. PLoS Genet, 2010. 6(9): p. e1001127.
103. Scuteri A, et al.: Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet, 2007. 3(7): p. e115.
104. Frayling TM, et al.: A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science, 2007. 316(5826): p. 889-94.
105. Loos RJ, et al.: Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nat Genet, 2008. 40(6): p. 768-75.
106. Willer CJ, et al.: Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet, 2009. 41(1): p. 25-34.
107. Thorleifsson G, et al.: Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet, 2009. 41(1): p. 18-24.

108. Kilpelainen TO, et al.: Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. Nat Genet, 2011. 43(8): p. 753-60.
109. Lindgren CM, et al.: Genome-Wide Association Scan Meta-Analysis Identifies Three Loci Influencing Adiposity and Fat Distribution. PLoS Genetics, 2009. 5(6).
110. Heard-Costa NL, et al.: NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. PLoS Genet, 2009. 5(6): p. e1000539.
111. Okada Y, et al.: Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet, 2012. 44(5): p. 511-6.
112. Wen W, et al.: Meta-analysis identifies common variants associated with body mass index in east Asians. Nat Genet, 2012. 44(3): p. 307-11.
113. Bradfield JP, et al.: A genome-wide association meta-analysis identifies new childhood obesity loci. Nat Genet, 2012.
114. Berndt SI, et al.: Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat Genet, 2013. 45(5): p. 501-12.
115. Fox CS, et al.: Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. PLoS Genet, 2012. 8(5): p. e1002695.
116. Guo Y, et al.: Gene-centric meta-analyses of 108 912 individuals confirm known body mass index loci and reveal three novel signals. Hum Mol Genet, 2013. 22(1): p. 184-201.
117. Randall JC, et al.: Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. PLoS Genet, 2013. 9(6): p. e1003500.
118. Melka MG, et al.: Genome-wide scan for loci of adolescent obesity and their relationship with blood pressure. J Clin Endocrinol Metab, 2012. 97(1): p. E145-50.
119. Paternoster L, et al.: Genome-wide population-based association study of extremely overweight young adults--the GOYA study. PLoS One, 2011. 6(9): p. e24303.
120. Scherag A, et al.: Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. PLoS Genet, 2010. 6(4): p. e1000916.
121. Jiao H, et al.: Genome wide association study identifies KCNMA1 contributing to human obesity. BMC Med Genomics, 2011. 4: p. 51.
122. Wang K, et al.: A genome-wide association study on obesity and obesity-related traits. PLoS One, 2011. 6(4): p. e18939.
123. Meyre D, et al.: Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. Nat Genet, 2009. 41(2): p. 157-9.
124. Cotsapas C, et al.: Common body mass index-associated variants confer risk of extreme obesity. Hum Mol Genet, 2009. 18(18): p. 3502-7.
125. Chambers JC, et al.: Common genetic variation near MC4R is associated with waist circumference and insulin resistance. Nat Genet, 2008. 40(6): p. 716-8.
126. Speliotes EK, et al.: Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet, 2011. 7(3): p. e1001324.
127. Kawaguchi T, et al.: Genetic polymorphisms of the human PNPLA3 gene are strongly associated with severity of non-alcoholic fatty liver disease in Japanese. PLoS One, 2012. 7(6): p. e38322.

128. DIAbetes Genetics Replication Meta-analysis Consortium, et al.: Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet, 2014. 46(3): p. 234-244.
129. ENCODE Project Consortium, et al.: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 2007. 447(7146): p. 799-816.
130. Gaulton KJ, et al.: A map of open chromatin in human pancreatic islets. Nat Genet, 2010. 42(3): p. 255-259.
131. Bernstein BE, et al.: An integrated encyclopedia of DNA elements in the human genome. Nature, 2012. 489(7414): p. 57-74.
132. Thurman RE, et al.: The accessible chromatin landscape of the human genome. Nature, 2012. 489(7414): p. 75-82.
133. da Silva Xavier G, et al.: TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells. Diabetes, 2009. 58(4): p. 894-905.
134. Boj Sylvia F, et al.: Diabetes Risk Gene and Wnt Effector Tcf7l2/TCF4 Controls Hepatic Response to Perinatal and Adult Metabolic Demand. Cell, 2012. 151(7): p. 1595-1607.
135. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, et al.: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science, 2007. 316(5829): p. 1331-6.
136. Sladek R, et al.: A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature, 2007. 445(7130): p. 881-5.
137. Beer NL, et al.: The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Hum Mol Genet, 2009. 18(21): p. 4081-8.
138. Small KS, et al.: Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet, 2011. 43(6): p. 561-4.
139. Prokopenko I, et al.: Variants in MTNR1B influence fasting glucose levels. Nat Genet, 2009. 41(1): p. 77-81.
140. Lyssenko V, et al.: Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat Genet, 2009. 41(1): p. 82-8.
141. Rung J, et al.: Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. Nat Genet, 2009. 41(10): p. 1110-5.
142. Freathy RM, et al.: Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI. Diabetes, 2008. 57(5): p. 1419-26.
143. Freathy RM, et al.: Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. Nat Genet, 2010. 42(5): p. 430-5.
144. Loos RJ and Yeo GS: The bigger picture of FTO--the first GWAS-identified obesity gene. Nat Rev Endocrinol, 2014. 10(1): p. 51-61.
145. Willems SM, et al.: A methodological perspective on genetic risk prediction studies in type 2 diabetes: recommendations for future research. Curr Diab Rep, 2011. 11(6): p. 511-8.
146. Walters RG, et al.: A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature, 2010. 463(7281): p. 671-5.
147. Lee HS, et al.: Genome-wide copy number variation study reveals KCNIP1 as a modulator of insulin secretion. Genomics, 2014. 104(2): p. 113-20.
148. Lee BY, et al.: Genome-wide analysis of copy number variations reveals that aging processes influence body fat distribution in Korea Associated Resource (KARE) cohorts. Hum Genet, 2012. 131(11): p. 1795-804.
149. Lupski JR, et al.: Clan genomics and the complex architecture of human disease. Cell, 2011. 147(1): p. 32-43.

150. Stahl EA, et al.: Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet, 2012. 44(5): p. 483-9.
151. Yang J, et al.: Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet, 2012. 44(4): p. 369-75, S1-3.
152. Zuk O, et al.: The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences, 2012.
153. Waddington C: The epigenotype. International Journal of Epidemiology, 2011.
154. Handel AE, Ebers GC, and Ramagopalan SV: Epigenetics: molecular mechanisms and implications for disease. Trends Mol Med, 2010. 16(1): p. 7-16.
155. Bannister AJ and Kouzarides T: Regulation of chromatin by histone modifications. Cell Res, 2011. 21(3): p. 381-395.
156. Li B, Carey M, and Workman JL: The Role of Chromatin during Transcription. Cell, 2007. 128(4): p. 707-719.
157. Kouzarides T: SnapShot: Histone-Modifying Enzymes. Cell, 2007. 131(4): p. 822-822.e1.
158. Matzke MA and Birchler JA: RNAi-mediated pathways in the nucleus. Nat Rev Genet, 2005. 6(1): p. 24-35.
159. Ji H, et al.: Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature, 2010. 467(7313): p. 338-342.
160. Curradi M, et al.: Molecular mechanisms of gene silencing mediated by DNA methylation. Mol Cell Biol, 2002. 22(9): p. 3157-73.
161. Medvedeva Y, et al.: Effects of cytosine methylation on transcription factor binding sites. BMC Genomics, 2014. 15(1): p. 119.
162. Estecio MR, et al.: SINE retrotransposons cause epigenetic reprogramming of adjacent gene promoters. Mol Cancer Res, 2012. 10(10): p. 1332-42.
163. Eckhardt F, et al.: DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet, 2006. 38(12): p. 1378-85.
164. Moore LD, Le T, and Fan G: DNA methylation and its basic function. Neuropsychopharmacology, 2012. 38(1): p. 23-38.
165. Li Y, et al.: The DNA Methylome of Human Peripheral Blood Mononuclear Cells. PLoS Biol, 2010. 8(11): p. e1000533.
166. Zaidi SK, et al.: Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information. Mol Cell Biol, 2010. 30(20): p. 4758-66.
167. Berger SL, et al.: An operational definition of epigenetics. Genes & development, 2009. 23(7): p. 781.
168. Morgan HD, et al.: Epigenetic reprogramming in mammals. Human Molecular Genetics, 2005. 14(suppl 1): p. R47-R58.
169. Zhang Y and Tycko B: Monoallelic expression of the human H19 gene. Nat Genet, 1992. 1(1): p. 40-4.
170. Huang RC, et al.: DNA methylation of the IGF2/H19 imprinting control region and adiposity distribution in young adults. Clin Epigenetics, 2012. 4(1): p. 21.
171. Richards EJ: Inherited epigenetic variation-revisiting soft inheritance. Nat Rev Genet, 2006. 7(5): p. 395-401.
172. Grossniklaus U, et al.: Transgenerational epigenetic inheritance: how important is it? Nat Rev Genet, 2013. 14(3): p. 228-235.
173. Ferguson-Smith AC and Patti M-E: You Are What Your Dad Ate. Cell Metabolism. 13(2): p. 115-117.
174. Morgan HD, et al.: Epigenetic inheritance at the agouti locus in the mouse. Nat Genet, 1999. 23(3): p. 314-8.
175. Eichten SR, et al.: Heritable Epigenetic Variation among Maize Inbreds. PLoS Genet, 2011. 7(11): p. e1002372.

176. Becker C, et al.: Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature, 2011. 480(7376): p. 245-249.
177. Dolinoy DC: The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. Nutr Rev, 2008. 66 Suppl 1: p. S7-11.
178. Gibbs JR, et al.: Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genetics, 2010. 6(5): p. e1000952.
179. Bell JT, et al.: DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol, 2011. 12(1): p. R10.
180. Shoemaker R, et al.: Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res, 2010. 20(7): p. 883-9.
181. Bell CG, et al.: Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. PLoS One, 2010. 5(11): p. e14040.
182. Seki Y, et al.: Minireview: Epigenetic programming of diabetes and obesity: animal models. Endocrinology, 2012. 153(3): p. 1031-8.
183. Lillycrop KA and Burdge GC: Epigenetic changes in early life and future risk of obesity. Int J Obes (Lond), 2010.
184. Burdge GC, et al.: Dietary protein restriction of pregnant rats in the F0 generation induces altered methylation of hepatic gene promoters in the adult male offspring in the F1 and F2 generations. Br J Nutr, 2007. 97(3): p. 435-9.
185. Kuehnen P, et al.: An Alu Element-Associated Hypermethylation Variant of the POMC Gene Is Associated with Childhood Obesity. PLoS Genet, 2012. 8(3): p. e1002543.
186. Herrera BM, Keildson S, and Lindgren CM: Genetics and epigenetics of obesity. Maturitas, 2011. 69(1): p. 41-9.
187. Milagro FI, et al.: A dual epigenomic approach for the search of obesity biomarkers: DNA methylation in relation to diet-induced weight loss. FASEB J, 2011. 25(4): p. 1378-89.
188. Dabelea D: The predisposition to obesity and diabetes in offspring of diabetic mothers. Diabetes Care, 2007. 30(Supplement 2): p. S169.
189. Dabelea D, et al.: Intrauterine exposure to diabetes conveys risks for type 2 diabetes and obesity: a study of discordant sibships. Diabetes, 2000. 49(12): p. 2208-2211.
190. Krude H, et al.: Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. Nat Genet, 1998. 19(2): p. 155-7.
191. Bibikova M, et al.: High density DNA methylation array with single CpG site resolution. Genomics, 2011. 98(4): p. 288-95.
192. Rakyan VK, et al.: Epigenome-wide association studies for common human diseases. Nat Rev Genet, 2011. 12(8): p. 529-41.
193. Toperoff G, et al.: Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. Human Molecular Genetics, 2012. 21(2): p. 371-383.
194. Almen MS, et al.: Genome wide analysis reveals association of a FTO gene variant with epigenetic changes. Genomics, 2012. 99(3): p. 132-7.
195. Liu ZH, et al.: Methylation Status of CpG Sites in the MCP-1 Promoter is Correlated to Serum MCP-1 in type 2 Diabetes. J Endocrinol Invest, 2011.
196. Yang L, et al.: Systematic evaluation of genome-wide methylated DNA enrichment using a CpG island array. BMC Genomics, 2011. 12: p. 10.
197. Yang BT, et al.: Increased DNA Methylation and Decreased Expression of PDX-1 in Pancreatic Islets from Patients with Type 2 Diabetes. Mol Endocrinol, 2012.

198. Ling C, et al.: Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia, 2008. 51(4): p. 615-22.
199. Barres R, et al.: Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. Cell Metab, 2009. 10(3): p. 189-98.
200. Volkmar M, et al.: DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. EMBO J, 2012. 31(6): p. 1405-26.
201. Dayeh TA, et al.: Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia, 2013. 56(5): p. 1036-1046.
202. Hidalgo B, et al.: Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. Diabetes, 2014. 63(2): p. 801-7.
203. Canivell S, et al.: Gastric inhibitory polypeptide receptor methylation in newly diagnosed, drug-naive patients with type 2 diabetes: a case-control study. PLoS One, 2013. 8(9): p. e75474.
204. Ma J, et al.: No association between IRS1 promoter methylation and type 2 diabetes. Mol Med Rep, 2013. 8(3): p. 949-53.
205. Relton CL, et al.: DNA methylation patterns in cord blood DNA and body size in childhood. PLoS One, 2012. 7(3): p. e31821.
206. Souren NYP, et al.: DNA Methylation Variability at Growth-Related Imprints Does not Contribute to Overweight in Monozygotic Twins Discordant for BMI. Obesity, 2011. 19(7): p. 1519-1522.
207. Stepanow S, et al.: Allele-specific, age-dependent and BMI-associated DNA methylation of human MCHR1. PLoS One, 2011. 6(5): p. e17711.
208. Godfrey KM, et al.: Epigenetic gene promoter methylation at birth is associated with child's later adiposity. Diabetes, 2011. 60(5): p. 1528-34.
209. Zhao J, Goldberg J, and Vaccarino V: Promoter methylation of serotonin transporter gene is associated with obesity measures: a monozygotic twin study. Int J Obes (Lond), 2012.
210. Groom A, et al.: Postnatal growth and DNA methylation are associated with differential gene expression of the TACSTD2 gene and childhood fat mass. Diabetes, 2012. 61(2): p. 391-400.
211. Feinberg AP, et al.: Personalized epigenomic signatures that are stable over time and covary with body mass index. Sci Transl Med, 2010. 2(49): p. 49ra67.
212. Wang X, et al.: Obesity related methylation changes in DNA of peripheral blood leukocytes. BMC Med, 2010. 8: p. 87.
213. Dick KJ, et al.: DNA methylation and body-mass index: a genome-wide analysis. Lancet, 2014. 383(9933): p. 1990-8.
214. Smyth G: Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor, 2005: p. 397-420.
215. Aryee MJ, et al.: Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics, 2014. 30(10): p. 1363-9.
216. Maksimovic J, Gordon L, and Oshlack A: SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol, 2012. 13(6): p. R44.
217. Dedeurwaerder S, et al.: Evaluation of the Infinium Methylation 450K technology. Epigenomics, 2011. 3(6): p. 771-84.
218. Teschendorff AE, et al.: A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics, 2013. 29(2): p. 189-96.
219. Touleimat N and Tost J: Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile

normalization for accurate DNA methylation estimation. Epigenomics, 2012. 4(3): p. 325-41.
220. Price ME, et al.: Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin, 2013. 6(1): p. 4.
221. Pinnick KE and Karpe F: DNA methylation of genes in adipose tissue. The Proceedings of the Nutrition Society, 2010: p. 1-7.
222. Walther TC and Farese RV, Jr.: Lipid droplets and cellular lipid metabolism. Annu Rev Biochem, 2012. 81: p. 687-714.
223. Porter SA, et al.: Abdominal subcutaneous adipose tissue: a protective fat depot? Diabetes Care, 2009. 32(6): p. 1068-75.
224. Tran TT, et al.: Beneficial effects of subcutaneous fat transplantation on metabolism. Cell Metab, 2008. 7(5): p. 410-20.
225. Gutierrez-Arcelus M, et al.: Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife, 2013. 2: p. e00523.
226. Kulis M, et al.: Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. Biochim Biophys Acta, 2013. 1829(11): p. 1161-74.
227. Kerkel K, et al.: Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet, 2008. 40(7): p. 904-8.
228. Schalkwyk LC, et al.: Allelic Skewing of DNA Methylation Is Widespread across the Genome. American journal of human genetics, 2010. 86(2): p. 196-212.
229. Zhang D, et al.: Genetic control of individual differences in gene-specific methylation in human brain. Am J Hum Genet, 2010. 86(3): p. 411-9.
230. Moen EL, et al.: Genome-Wide Variation of Cytosine Modifications Between European and African Populations and the Implications for Complex Traits. Genetics, 2013.
231. McRae AF, et al.: Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biol, 2014. 15(5): p. R73.
232. Grundberg E, et al.: Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet, 2013. 93(5): p. 876-90.
233. Tan GD, et al.: The in vivo effects of the Pro12Ala PPARgamma2 polymorphism on adipose tissue NEFA metabolism: the first use of the Oxford Biobank. Diabetologia, 2006. 49(1): p. 158-68.
234. Huang TH, Perry MR, and Laux DE: Methylation profiling of CpG islands in human breast cancer cells. Hum Mol Genet, 1999. 8(3): p. 459-70.
235. Fassbender A, et al.: Quantitative DNA methylation profiling on a high-density oligonucleotide microarray. Methods Mol Biol, 2010. 576: p. 155-70.
236. Schumacher A, et al.: Microarray-based DNA methylation profiling: technology and applications. Nucleic Acids Res, 2006. 34(2): p. 528-42.
237. Lewin J, et al.: Comparative DNA methylation analysis in normal and tumour tissues and in cancer cell lines using differential methylation hybridisation. Int J Biochem Cell Biol, 2007. 39(7-8): p. 1539-50.
238. Durbin RM, et al.: A map of human genome variation from population-scale sequencing. Nature, 2010. 467(7319): p. 1061-1073.
239. Howie BN, Donnelly P, and Marchini J: A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLoS Genet, 2009. 5(6): p. e1000529.
240. Dai M, et al.: Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res, 2005. 33(20): p. e175.
241. Irizarry RA, et al.: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics, 2003. 4(2): p. 249-64.

242. Nica AC, et al.: The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet, 2011. 7(2): p. e1002003.
243. Grundberg E, et al.: Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet, 2012. 44(10): p. 1084-9.
244. Spector TD and Williams FM: The UK Adult Twin Registry (TwinsUK). Twin Res Hum Genet, 2006. 9(6): p. 899-906.
245. Andrew T, et al.: Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. Twin Res, 2001. 4(6): p. 464-77.
246. Teschendorff AE, et al.: An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One, 2009. 4(12): p. e8274.
247. R Devlopment Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna Austria ISBN, 2011. 3(10).
248. Rossin EJ, et al.: Proteins Encoded in Genomic Regions Associated with Immune-Mediated Disease Physically Interact and Suggest Underlying Biology. PLoS Genet, 2011. 7(1): p. e1001273.
249. Ashburner M, et al.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet, 2000. 25(1): p. 25-9.
250. Morison IM, Paton CJ, and Cleverley SD: The imprinted gene and parent-of-origin effect database. Nucleic Acids Research, 2001. 29(1): p. 275.
251. Zeggini E, et al.: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet, 2008. 40(5): p. 638-645.
252. Bell JT, et al.: Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet, 2012. 8(4): p. e1002629.
253. Howie B, et al.: Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet, 2012. 44(8): p. 955-9.
254. Boyce BF and Xing L: Biology of RANK, RANKL, and osteoprotegerin. Arthritis Res Ther, 2007. 9 Suppl 1: p. S1.
255. Chang YH, et al.: Serum osteoprotegerin and tumor necrosis factor related apoptosis inducing-ligand (TRAIL) are elevated in type 2 diabetic patients with albuminuria and serum osteoprotegerin is independently associated with the severity of diabetic nephropathy. Metabolism, 2011. 60(8): p. 1064-9.
256. Reinhard H, et al.: Osteoprotegerin and mortality in type 2 diabetic patients. Diabetes Care, 2010. 33(12): p. 2561-6.
257. Nabipour I, et al.: Osteoprotegerin in relation to type 2 diabetes mellitus and the metabolic syndrome in postmenopausal women. Metabolism, 2010. 59(5): p. 742-7.
258. Panteghini M: Aspartate aminotransferase isoenzymes. Clin Biochem, 1990. 23(4): p. 311-9.
259. Aggerbeck M, et al.: Regulation of the cytosolic aspartate aminotransferase housekeeping gene promoter by glucocorticoids, cAMP, and insulin. Biochemistry, 1993. 32(35): p. 9065-72.
260. Beck S: Taking the measure of the methylome. Nat Biotechnol, 2010. 28(10): p. 1026-8.
261. Bock C: Analysing and interpreting DNA methylation data. Nat Rev Genet, 2012. 13(10): p. 705-719.
262. Liu Y, et al.: Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol, 2013. 31(2): p. 142-7.
263. Armstrong DA, et al.: Global and gene-specific DNA methylation across multiple tissues in early infancy: implications for children's health research. FASEB J, 2014. 28(5): p. 2088-97.

264. Horvath S, et al.: Aging effects on DNA methylation modules in human brain and blood tissue. Genome Biology, 2012. 13(10): p. R97.
265. Aberg KA, et al.: Testing two models describing how methylome-wide studies in blood are informative for psychiatric conditions. Epigenomics, 2013. 5(4): p. 367-77.
266. Smith ZD, et al.: DNA methylation dynamics of the human preimplantation embryo. Nature, 2014. 511(7511): p. 611-615.
267. Martino D, et al.: Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. Genome Biology, 2013. 14(5): p. R42.
268. Weisberg SP, et al.: Obesity is associated with macrophage accumulation in adipose tissue. J Clin Invest, 2003. 112(12): p. 1796-1808.
269. Martinez JA, et al.: Interplay of early-life nutritional programming on obesity, inflammation and epigenetic outcomes. Proc Nutr Soc, 2012: p. 1-8.
270. Houseman EA, et al.: DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics, 2012. 13(1): p. 86.
271. Rönn T, et al.: A Six Months Exercise Intervention Influences the Genome-wide DNA Methylation Pattern in Human Adipose Tissue. PLoS Genet, 2013. 9(6): p. e1003572.
272. Bibikova M, et al.: Genome-wide DNA methylation profiling using Infinium(R) assay. Epigenomics, 2009. 1(1): p. 177-200.
273. Bergman Y and Cedar H: DNA methylation dynamics in health and disease. Nat Struct Mol Biol, 2013. 20(3): p. 274-81.
274. Irizarry RA, et al.: The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet, 2009. 41(2): p. 178-186.
275. Stevens M, et al.: Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. Genome Res, 2013. 23(9): p. 1541-53.
276. Jaffe AE, et al.: Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. International Journal of Epidemiology, 2012. 41(1): p. 200-209.
277. Ong M-L and Holbrook JD: Novel region discovery method for Infinium 450K DNA methylation data reveals changes associated with aging in muscle and neuronal pathways. Aging Cell, 2014. 13(1): p. 142-155.
278. Morris TJ, et al.: ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics, 2014. 30(3): p. 428-30.
279. Sandoval J, et al.: Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics, 2011. 6(6): p. 692-702.
280. Illumina. *Illumina Genome Studio*. 2013; Available from: http://res.illumina.com.
281. Wang D, et al.: IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics, 2012. 28(5): p. 729-30.
282. Teschendorff AE, Zhuang J, and Widschwendter M: Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics, 2011. 27(11): p. 1496-505.
283. Sun Z, et al.: Batch effect correction for genome-wide methylation data with Illumina Infinium platform. BMC Med Genomics, 2011. 4: p. 84.
284. Zykovich A, et al.: Genome-wide DNA methylation changes with age in disease-free human skeletal muscle. Aging Cell, 2014. 13(2): p. 360-366.
285. Horvath S: DNA methylation age of human tissues and cell types. Genome Biol, 2013. 14(10): p. R115.

286. Vozarova B, et al.: High White Blood Cell Count Is Associated With a Worsening of Insulin Sensitivity and Predicts the Development of Type 2 Diabetes. Diabetes, 2002. 51(2): p. 455-461.
287. Laurent L, et al.: Dynamic changes in the human methylome during differentiation. Genome Res, 2010. 20(3): p. 320-31.
288. Lister R, et al.: Human DNA methylomes at base resolution show widespread epigenomic differences. Nature, 2009. 462(7271): p. 315-22.
289. Horton R, et al.: Variation analysis and gene annotation of eight MHC haplotypes: The MHC Haplotype Project. Immunogenetics, 2008. 60(1): p. 1-18.
290. Reinius LE, et al.: Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One, 2012. 7(7): p. e41361.
291. Du P, Kibbe WA, and Lin SM: lumi: a pipeline for processing Illumina microarray. Bioinformatics, 2008. 24(13): p. 1547-8.
292. Chambers JC, et al.: Plasma homocysteine concentrations and risk of coronary heart disease in UK Indian Asian and European men. Lancet, 2000. 355(9203): p. 523-7.
293. Relton CL and Davey Smith G: Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. International Journal of Epidemiology, 2012. 41(1): p. 161-176.
294. Chanock SJ, et al.: Replicating genotype-phenotype associations. Nature, 2007. 447(7145): p. 655-60.
295. Phillips T: The role of methylation in gene expression. Nature Education, 2008. 1(1): p. 116.
296. Lowel H, et al.: [Epidemiology of hypertension in Germany. Selected results of population-representative cross-sectional studies]. Dtsch Med Wochenschr, 2006. 131(46): p. 2586-91.
297. Devlin B and Roeder K: Genomic control for association studies. Biometrics, 1999. 55(4): p. 997-1004.
298. Schurmann C, et al.: Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. PloS one, 2012. 7(12): p. e50938.
299. Slieker RC, et al.: Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics Chromatin, 2013. 6(1): p. 26.
300. Liu Y, et al.: Methylomics of gene expression in human monocytes. Hum Mol Genet, 2013. 22(24): p. 5065-74.
301. Flicek P, et al.: Ensembl 2014. Nucleic Acids Research, 2014. 42(D1): p. D749-D755.
302. Segre AV, et al.: Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet, 2010. 6(8).
303. Chahal NS, et al.: Ethnicity-related differences in left ventricular function, structure and geometry: a population study of UK Indian Asian and European white subjects. Heart, 2010. 96(6): p. 466-71.
304. ENCODE Project Consortium: A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol, 2011. 9(4): p. e1001046.
305. Ram O, et al.: Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. Cell, 2011. 147(7): p. 1628-39.
306. Hellman A and Chess A: Extensive sequence-influenced DNA methylation polymorphism in the human genome. Epigenetics Chromatin, 2010. 3(1): p. 11.

307. Gertz J, et al.: Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. PLoS Genet, 2011. 7(8): p. e1002228.
308. Jia D, et al.: Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature, 2007. 449(7159): p. 248-51.
309. Suzuki MM and Bird A: DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet, 2008. 9(6): p. 465-76.
310. Deaton AM, et al.: Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res, 2011. 21(7): p. 1074-86.
311. Maunakea AK, et al.: Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature, 2010. 466(7303): p. 253-7.
312. Deaton AM and Bird A: CpG islands and the regulation of transcription. Genes Dev, 2011. 25(10): p. 1010-22.
313. Sigurdsson MI, et al.: HapMap methylation-associated SNPs, markers of germline DNA methylation, positively correlate with regional levels of human meiotic recombination. Genome Res, 2009. 19(4): p. 581-9.
314. Ling C, et al.: Genetic and epigenetic factors are associated with expression of respiratory chain component NDUFB6 in human skeletal muscle. J Clin Invest, 2007. 117(11): p. 3427-35.
315. Bird A: The dinucleotide CG as a genomic signalling module. J Mol Biol, 2011. 409(1): p. 47-53.
316. Cooper DN and Krawczak M: The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. Hum Genet, 1990. 85(1): p. 55-74.
317. Zhi D, et al.: SNPs located at CpG sites modulate genome-epigenome interaction. Epigenetics, 2013. 8(8): p. 802-6.
318. Gabriel S: Variation in the human genome and the inherited basis of common disease. Semin Oncol, 2006. 33(6 Suppl 11): p. S46-9.
319. Rosas-Vargas H, Martinez-Ezquerro JD, and Bienvenu T: Brain-derived neurotrophic factor, food intake regulation, and obesity. Arch Med Res, 2011. 42(6): p. 482-94.
320. Brown CD, Mangravite LM, and Engelhardt BE: Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. PLoS Genet, 2013. 9(8): p. e1003649.
321. Manor O and Segal E: Robust prediction of expression differences among human individuals using only genotype information. PLoS Genet, 2013. 9(3): p. e1003396.
322. Bae JB: Perspectives of international human epigenome consortium. Genomics Inform, 2013. 11(1): p. 7-14.
323. Rakyan VK, et al.: Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. PLoS Genet, 2011. 7(9): p. e1002300.
324. Pidsley R, et al.: A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics, 2013. 14: p. 293.
325. Wu MC, et al.: A systematic assessment of normalization approaches for the Infinium 450K methylation platform. Epigenetics, 2014. 9(2): p. 318-29.
326. Jaffe AE and Irizarry RA: Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol, 2014. 15(2): p. R31.
327. Shoemaker R, Wang W, and Zhang K: Mediators and dynamics of DNA methylation. Wiley Interdiscip Rev Syst Biol Med, 2011. 3(3): p. 281-98.
328. Jaffe AE and Kaminsky Z: FlowSorted.DLPFC.450k: Illumina HumanMethylation data on sorted frontal cortex cell populations. R package version 1.0.0.

329. Smith GD, et al.: Clustered Environments and Randomized Genes: A Fundamental Distinction between Conventional and Genetic Epidemiology. PLoS Med, 2007. 4(12): p. e352.

Published manuscripts