DNA methylation signature in type 1 diabetes

Mary Anh Ngoc Dang

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

Supervisors
Professor R.D.G. Leslie
Professor P. Pozzilli

Department of Endocrinology and Diabetes
Università Campus Bio-Medico di Roma
Via Alvaro del Portillo
21-00128 Roma
Italy

January 2015

Statement of Originality

I, Mary Anh Ngoc Dang, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: Mary Anh Ngoc Dang
Date: January 2015

Details of collaboration and publications:

Professor R. David Leslie	Lead of The British Diabetic Twin Study
Dr Mohammed I. Hawa	Performed radioimmunoassay; Section 2.2
Stephanie Cunningham	Recruitment of twin pairs
Professor Vardhman Rakyan	Provided DNA in Section 4.3.9 BS-seq library preparation; Section 2.11
Dr Robert Lowe	Illumina450K, BS-seq and power analyses; Section 3.3.1, 3.3.2, 3.3.3, 3.3.7, 3.3.8; Figures 3.1A and B, 3.2, 3.3, 3.7, 3.8, 3.9, 3.10
Dr Dirk S. Paul	Illumina450K analysis; Section 3.1, 3.4, 3.3.5; Figures 3.1C, 3.4, 3.5, 3.6A
Dr Emanuele Libertini	BS-seq analysis; Section 3.3.9; Figures 3.11; Table 3.2
Professor Ake Lernmark	Provided cord blood samples
Rasmus Bennet	Provided cord blood samples
UCL Genome Centre, London	Illumina450K array; Section 2.10.1.1
Centro Nacional de Análisis Genómico (CNAG), Barcelona	BS-Seq; Section 2.11.5
Barts and the London Genome Centre, London	DNA fingerprinting and pyrosequencing; Section 2.1, 2.14

Abstract

Type 1 diabetes is an autoimmune disease due to the interaction of genetic and non-genetic factors, leading to an immune response against insulin secreting islet cells. Concordance rates for type 1 diabetes in monozygotic twins vary widely and no single environmental factor has been shown to cause the disease. Therefore, epigenetics has been suggested to play a role in diabetes aetiology. Preliminary results identified DNA methylation changes in $CD14^+$ monocytes from childhood-onset type 1 diabetes which antedated the disease.

Following on from this work, this present study was carried out to investigate whole-genome DNA methylation profiles in $CD14^+CD16^-$ monocytes, $CD4^+$ T cells, $CD19^+$ B cells and buccal cells from 24 monozygotic twin pairs discordant for type 1 diabetes. DNA methylation was profiled using Illumina Infinium HumanMethylation450K BeadChip and analysed using the ChAMP pipeline. Bisulfite sequencing was also carried out on $CD4^+$ cells from four monozygotic twin pairs also discordant for type 1 diabetes. Through bioinformatics analyses, 258 cell-type specific differentially-methylated positions were identified from the 450K BeadChip and 125 differentially-methylated regions from bisulfite sequencing. DNA methylation was also shown to be stable, as similar methylation differences found in the preliminary study, were again detected in the same twin pairs sampled years later. As DNA methylation is a stable marker, it could be used as a biomarker. β-cell death in diabetes releases DNA with unmethylated CpG sites in the insulin promoter region into the blood circulation. To detect these differences, an assay was also developed testing serum samples from monozygotic twin pairs.

The data presented provided comprehensive DNA methylation profiles in type 1 diabetes from this discovery cohort. The methylation signature found will then be validated in diabetic, pre-diabetic and control singletons. This in turn will provide data for later functional analyses to identify genes associated with type 1 diabetes risk.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 8

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 8

Abbreviations and Units

Abbreviation	
(1,25(OH)2D)	1,25-dihydroxyvitamin D
3'UTR	3' untranslated region
5'UTR	5' untranslated region
5-hmC	5-hydroxymethylcytosine
5-mC	5-methylcytosine
25(OH)D	25-hydroxyvitamin D
aDMRs	Ageing-associated differentially methylated regions
AGEs	Advanced glycation end products
APCs	Antigen-presenting cells
APS	Adenosine 5' phosphosulfate
ATP	Adenosine triphosphate
bp	Base pair
BCRs	B cell receptors
BSP	Bisulfite-specific PCR
BS-seq	Bisulfite sequencing
CD	Cluster of differentiation
cDNA	Complementary DNA
CG	Cytosine-guanine
CGIs	CpG islands
ChAMP	Chip Analysis Methylation Pipeline
ChIP-seq	Chromatin Immunoprecipitation sequencing
CpG	Cytosine-phosphate-guanine
Cq	Quantification cycle
CTLA4	Cytotoxic T lymphocyte antigen 4
CXCL2	Chemokine (C-X-C motif) ligand 2
DASP	The Diabetes Antibody Standardization Program
DAVID	Database for Annotation, Visualization and Integrated Discovery
DBS	Dried blood spots
ddPCR	Droplet Digital™ PCR
DHSs	DNase hypersensitivity sites
DIAMOND	Multinational Project for Childhood Diabetes

DMPs	Differentially-methylated positions
DMSO	Dimethyl sulfoxide
DMRs	Differentially methylated regions
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNMTs	DNA methyltransferases
dNTP	Deoxyribonucleotide triphosphate
DZ	Dizygotic
EBV	Epstein-Barr Virus
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ENCODE	Encyclopedia of DNA Elements
EURODIAB	Epidemiology and Prevention of Diabetes
EWAS	Epigenome-wide association studies
FACS	Fluorescence-activated cell sorting
FFPE	Formalin-foxed paraffin-embedded
FMO	Fluoroscence minus one
FOXP3	Forkhead box P3
FRET	Fluorescence resonance energy transfer
GAD	Glutamic acid decarboxylase
GADA	Glutamic acid decarboxylase autoantibodies
GWAS	Genome-wide association studies
HATs	Histone acetyltransferases
HDACs	Histone deacetylases
HLA	Human leukocyte antigen
IA-2	Insulinoma Antigen 2
IA-2A	Insulinoma Antigen 2 autoantibodies
IAA	Insulin autoantibodies
IDAT	Intensity Data files
IDDM2	Insulin-Dependent Diabetes Mellitus 2
IDIN	IRF7-driven inflammatory network
IFIH1	Interferon induced with helicase C domain 1
IFN	Interferon
IgG1	Immunoglobulin G1
IHEC	International Human Epigenome Consortium
IL-1	Interleukin-1

Illumina450K	Illumina Infinium HumanMethylationBeadChip450K
INS	Insulin gene
IRF7	Interferon regulatory factor 7
kb	Kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LADA	Latent autoimmune diabetes of adults
LADs	Lamina associated domains
LB	Lysogeny broth
LED	Light-emitting diode
lincRNAs	Large intergenic non-coding RNAs
LINE-1	Long interspersed element-1
LPS	Lipopolysaccharide
M bead	Methylated bead
MACS	Magnetic Activated Cell Sorting
MBD	Methyl CpG binding domain
MDA5	Melanoma differentiation-associated protein 5
MeDIP	Methylated DNA immunoprecipitation
MHC	Major histocompatibility complex
miRNAs	MicroRNAs
mPOD	Methylation profile of DNA
mRNA	Messenger RNA
MS	Multiple sclerosis
MSP	Methylation-specific PCR
MVPs	Methylation variable positions
MZ	Monozygotic
ncRNAs	Non-coding RNAs
NOD	Non-obese diabetic
NTC	No-template control
PAMPs	Pathogen-associated molecular patterns
PBMCs	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
piRNAs	PIWI-interacting RNAs
PPi	Pyrophosphate
PPI	Preproinsulin

PPP1R1A	Protein phosphatase 1, regulatory (inhibitor) subunit 1A
PRRs	Pattern recognition receptors
qPCR	Quantitative real-time PCR
qRT-PCR	Quantitative reverse transcription PCR
RA	Rheumatoid arthritis
REC	Research Ethics Committee
RNA	Ribonucleic acid
rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute
RRBS	Reduced representation bisulphite sequencing
RT	Room temperature
SEARCH	SEARCH for Diabetes in Youth Study
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SP	Sequencing primer
ssDNA	Single-stranded DNA
T1D	Type 1 diabetes mellitus
T1DGC	The International Type 1 Diabetes Genetics Consortium
T1D-MVPs	T1D–associated methylation variable positions
T2D	Type 2 diabetes mellitus
TBE	Tris/Borate/EDTA
T_{CM}	Central memory T cells
TCRs	T-cell receptors
tDMRs	Tissue-specific DMRs
TE	Tris-EDTA
T_{EM}	Effector memory T cells
TET	Ten-eleven translocation
TF	Transcription factor
TFBS	Transcription factor binding site
T_H cells	T helper cells
TLRs	Toll-like receptors
T_m	Melting temperature
TNF	Tumour necrosis factor
Tregs	Regulatory T cells
Tris	tris(hydroxymethyl)aminomethane
TTBS	Tris Buffered Saline tween

U bead	Unmethylated bead
VNTR	Variable number tandem repeat
WHO	World Health Organisation
WTCCC	The Wellcome Trust Case Control Consortium
ZnT8	Zinc transporter 8
ZnT8A	Zinc transporter 8 antibody

Units	
g	Gram
L	Litre
M	Molar (mole/litre)
mg	Milligram
mL	Millilitre
ng	Nanogram
nM	Nanomolar
pH	Log [H+]
SD	Standard deviation
µL	Microlitre
%	Percentage
±	Standard deviation
°C	Degrees centigrade
xg	Centrifugal force

Published Abstracts

1. Bradford CM, Hawa M, Sumaria N, **Dang MN**, Beyan H, Davies G, H. Jensen H, Tsang V, Pennington DJ & Leslie RD (2013) Understanding auto-antigen expression and regulation in human medullary thymic epithelial cells (mTECs). *Immunology*. 140:187, Suppl. 1. British Society for Immunology Congress 2013, Liverpool, UK

Published Papers

The following papers have been published or in preparation in support of this thesis.

1. **Dang MN**, Buzzetti R, Pozzilli P. (2013) Epigenetics in autoimmune diseases with focus on type 1 diabetes. *Diabetes Metab Res Rev*. 29(1):8-18.

Acknowledgements

There are a number of people I would like to thank for getting me through this PhD.

First I would like to express my gratitude to my supervisors, Professor David Leslie and Professor Paolo Pozzilli, for the opportunity to embark on this journey to a PhD. Their knowledge, guidance and patience throughout have given me invaluable skills for the future. Thank you to Campus Biomedico and the Leslie lab for funding this PhD.

I am also thankful to Dr Mohammed Hawa, Stephanie Cunningham, Dr Lina Paschou, Dr Claire Bradford and Dr Huriya Beyan for their support and for sharing their experiences relating to the lab. I express warm thanks to Capucine Grandjean and Dr William Ogunkolade for their advice on the qPCR experiments and I would like to thank former and current colleagues in the Diabetes department for the great atmosphere in the lab, particularly when working on the weekends or late nights.

I am grateful to the colleagues that are part of the BLUEPRINT consortium. Thank you to Professor Willem Ouwehand, Dr Kate Downes and Dr Mattia Frontini in Cambridge for getting me started with this project. Thanks to Professor Ake Lernmark and Rasmus Bennet in Sweden for providing cord blood samples and to Professor Stephan Beck and Professor Vardhman Rakyan in London for sharing their expertise.

Special thanks to Dr Rob Lowe, Dr Dirk Paul and Dr Emanuele Libertini for their help and guidance in statistical support and data analyses and therefore contributing to my thesis.

Thank you to my colleagues in Rome, for the support and hospitality they have shown with my every visit to Rome. Thank you to all the twin pairs that have volunteered their time and effort to take part in this study.

Last but not least, I would like to thank my family and friends for their support and encouragement during this PhD.

Chapter 1

General introduction

1 General introduction

Epigenetic mechanisms have been studied in diseases such as systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) (Dang *et al.*, 2013; Lu, 2013) and cancer (Rodriguez-Paredes and Esteller, 2011), however, there is little research focusing on type 1 diabetes mellitus (T1D). This thesis will centre on identifying DNA methylation differences in monozygotic (MZ) twin pairs discordant for T1D with different methylation profiling technologies such as the Illumina Infinium HumanMethylation450K BeadChip and bisulfite sequencing. The following introduction focuses on the general aspects of autoimmunity, T1D and epigenetics.

1.1 The immune system

The immune system is complex and has evolved to defend the body from foreign pathogens. This system maintains homeostasis, which in turn avoids chronic inflammatory processes and autoimmune disease. The first lines of defence are the anatomic barriers such as skin, tears and saliva which function to prevent pathogens from entering the body (Turvey and Broide, 2010). However, some pathogens are still able to successfully enter the body. These pathogens would then encounter the innate immune system. Another subsystem is the adaptive immune system, which is antigen-specific. These different systems differ in the recognition of triggering factors and their ability to respond to these signals.

1.1.1 Innate immunity

Innate immunity refers to the rapid self-protection from foreign pathogens by discriminating between host cells and pathogens. This system acts quickly, responding within minutes or hours of exposure to the pathogen. The response that follows involves the recruitment of certain cell populations to engulf and kill pathogens at the site of infection. Haematopoietic cells involved include monocytes, macrophages, neutrophils, natural killer cells and eosinophils (Turvey and Broide, 2010). Pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs) are expressed in the innate immune system to detect pathogen-associated molecular patterns (PAMPs) expressed by bacteria and viruses such as lipopolysaccharide (LPS), the major component of the outer wall of Gram-negative bacteria (Medzhitov and Janeway, 2002). Once a foreign pathogen has been detected, this activates immune cells such as monocytes to release stimulatory factors to begin the process of eliminating the infection (Medzhitov, 2001; Medzhitov, 2007).

1.1.1.1 Monocytes and macrophages

Monocytes are necessary components in innate immunity and are released into the blood from the bone marrow. Monocytes are a heterogeneous cell population in which there are two main classes; classical monocytes ($CD14^+CD16^-$) which represent approximately 90-95% of total blood monocytes and the non-classical monocytes ($CD14^+CD16^+$) (Parihar *et al.*, 2010; Gordon and Taylor, 2005). Aside from antigen presentation, differences between the subsets include cytokine production, however, exactly what cytokines are produced by which subset is still yet to be defined (Wong *et al.*, 2012). The non-classical subset is also more susceptible to apoptosis induced by reactive oxygen species (Zhao *et al.*, 2010). TLR4 is the best characterised TLR and once LPS binds to TLR4, monocytes release inflammatory cytokines such as tumour necrosis factor (TNF) and interleukin-12 (IL-12) (Saha and Geissmann, 2011). These activating cytokines stimulate the cells to eliminate noxious stimuli. Upon inflammation, circulating monocytes which do not undergo apoptosis differentiate into macrophages in tissues. Macrophages are phagocytic cells which have pathogen-recognition receptors and induce the production of inflammatory cytokines (Geissmann *et al.*, 2010). Macrophages have longer life spans compared to monocytes which remain in the blood for up to three days before spontaneously undergoing apoptosis in order to regulate the immune cells (Gonzalez-Mejia and Doseff, 2009). Monocytes and macrophages are also involved in the activation of the adaptive immune response by stimulating T cell priming (Geissmann *et al.*, 2008).

1.1.2 Adaptive immunity

The adaptive immune system responds less rapidly but creates an immunological memory from a primary response to a specific pathogen. Here, the adaptive component of the immune system generates an efficient immune response after three to five days against that specific pathogen and then more rapidly once the body is exposed to that pathogen again (Schenten and Medzhitov, 2011). The adaptive immune response uses T and B cells, with their respective receptor, to recognise potentially dangerous molecules. B cells are matured in the bone marrow and produce antibodies whereas the T cells are matured in the thymus and both are involved in the humoral immune response and cell-mediated immune response respectively (Zhao *et al.*, 2012).

1.1.2.1 B cells

B cells express cell surface immunoglobulin receptors that recognises specific antigenic epitopes (LeBien and Tedder, 2008) and accounts for 5-15% of circulating lymphocytes (Fettke *et al.*, 2014). The main function of B cells is the production of antibodies against microbial antigens (Fettke *et al.*, 2014). An antibody has two identical light and two identical heavy chains and the functional rearrangement of the Ig loci involves three segments: V, D and J in the heavy-chain variable region. These segments recombine randomly, in a process called VDJ recombination, to produce a unique variable domain in the immunoglobulin of each B cell (LeBien and Tedder, 2008). B cells develop in the bone marrow in several stages, differentiating from haematopoietic stem cells before migrating to secondary lymphoid tissues (Ichii *et al.*, 2014). Pro-B cells are the earliest committed precursor of the B cell lineage. These pro-B cells become pre- B cells, leading to immature B cells and then mature B cells (Pieper *et al.*, 2013). All B cells express CD19, even at the early stages of B cell development (LeBien and Tedder, 2008). Together with CD21, CD19 forms a protein complex which is essential for B cell activation (Depoil *et al.*, 2008).

Following cell maturation, B cells remain in peripheral tissues until they encounter an antigen. B cells can undergo T-cell independent antigen activation or T-cell dependent antigen activation (LeBien and Tedder, 2008). T-cell independent antigen activation involves B cell receptors (BCRs) directly binding to thymus-independent antigens which directly activates B cells in order to proliferate into immunoglobulin-secreting cells. These antigens are normally of microbial origin such as LPS from Gram-negative bacteria. T-cell dependent antigen activation of B cells involves an antigen binding to BCRs which is then internalised and processed intracellularly. The antigenic peptides are bound to a major histocompatibility complex (MHC) II molecule and then presented on the B cell surface acting as an antigen-presenting cell (APC) to T cells (Wong *et al.*, 2004). The activated T cell then secretes cytokines which go on to activate the B cell. Eventually upon activation, most B cells proliferate and differentiate into antibody-secreting plasma cells or memory B cells. Subsequent exposure to the same antigen leads to a greater antibody response through the production of large quantities of high-affinity, mono-specific class-switched IgG antibodies. Naïve B cells, co-expressing IgM and IgD, exit the bone marrow and migrate to the spleen, where they divide into further three subsets: B1 cells, follicular B cells and marginal zone B cells (Allman and Pillai, 2008).

1.1.2.2 T cells

The development of T cells begins in the thymus, starting with a population of immature thymocytes (Naito *et al.*, 2011). The thymocytes are classed as double-negative as they do not express CD4 or CD8. As they progress, they become double-positive, $CD4^+CD8^+$. Double-positive thymocytes will then bind with MHC class I or MHC class II molecules to develop into $CD8^+$ or $CD4^+$ T cells respectively (Germain, 2002). These cells are then released from the thymus to peripheral tissues. Any cell that responds to the self-peptide-self MHC ligands undergoes negative selection, resulting in apoptosis. In addition, any cells that do not respond to the self-peptide-self-MHC ligand then undergo delayed apoptosis (Janeway, 2001). Activating an immune response in naïve T cells, involves an interaction between the APCs, which displays antigens bound to MHC class I or class II molecules on its surface and presents them to the T cells (Naito *et al.*, 2011). T cell receptors (TCRs) are integral membrane proteins and when coupled with CD3, induce a network of downstream signalling pathways, which eventually lead to cell proliferation and differentiation into specific effector cells. (Luckheeram *et al.*, 2012).

There are different subsets of $CD4^+$ T cells, the majority of which are helper cells (T_H cells) (Mueller *et al.*, 2013). T_H cells are involved in the activation of cytotoxic T cells and macrophages and assist in the maturation of B cells into plasma cells and memory B cells. Once T_H cells are activated, they can divide into further subsets which include T_H1, T_H2 or T_H17 cells which secrete different cytokines such as TNF-α, IL-4 and IL-12 respectively (Farber *et al.*, 2014). Memory T cells remain in the immune system long after an infection has resolved contributing to the life-long immunity that protects the body (Farber *et al.*, 2014). Upon re-exposure to a previously encountered antigen, memory T cells mount a response to produce large numbers of effector T cells. Memory T cells are further divided into subsets, central memory (T_{CM}) T cells and effector memory (T_{EM}) T cells, which are differentiated by the expression of chemokine receptors CCR7 and CD62L (Farber *et al.*, 2014).

Regulatory T cells (T_{regs}) are involved in the maintenance of immunological tolerance. Their main role is to down-modulate an immune response in order to prevent autoimmune diseases by supressing autoreactive T cells that escaped negative selection in the thymus (Sakaguchi *et al.*, 2008). The main markers for T_{regs} are forkhead box P3 (FOXP3), CD25, CTLA4 and CD127 (Rowe *et al.*, 2012). FOXP3 is required for the T_{regs} suppressive function (Larosa and Orange, 2008).

Function mutations in the FOXP3 gene result in the inflammatory autoimmune disease immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) (Bacchetta *et al.*, 2006).

1.2 Autoimmunity

Normally the immune system protects the host against foreign pathogens through self-tolerance. However, the body can elicit an immune response targeted at an individual's own cells and tissues when self-reactive lymphocytes elude tolerance. This process is called autoimmunity and it can cause inflammation and damage (Bach, 2005; Goodnow *et al.*, 2005). A disease can be defined as an autoimmune disease based on the following: there is direct evidence from transfer of pathogenic antibody or pathogenic T cells; there is indirect evidence based on reproduction of the autoimmune disease in experimental animals and there is circumstantial evidence from clinical clues (Rose and Bona, 1993). Examples of autoimmune diseases include SLE, RA, Graves' disease and T1D.

1.2.1 Genetic factors in autoimmunity

Disease concordance between identical twins indicates that genetic factors play a role in the prevalence of autoimmune diseases (Hewagama and Richardson, 2009). The MHC is the strongest association in most autoimmune diseases (Gregersen and Behrens, 2006). Many publications on genome wide association studies (GWAS) have identified other common variants which confer risk or protection in several autoimmune diseases (Baranzini, 2009; Lettre and Rioux, 2008). Some genes are associated with one or more disease (Zenewicz *et al.*, 2010; Rioux and Abbas, 2005). For example, *PTPN22*, which is involved in T and B cell receptor signalling, is a risk factor for RA and T1D. Another is *CTLA4*, an inhibitory receptor expressed by T cells and is associated with RA and T1D (Kormendy *et al.*, 2013; Chen *et al.*, 2013). Autoimmune diseases such as coeliac disease, autoimmune thyroid disease and RA, share multiple susceptibility genes (Parkes *et al.*, 2013). However, the same SNP in a number of loci shared between the autoimmune diseases, shows the strongest association, but in the opposite direction. The authors also reported that the risk alleles conferring the largest effect sizes were usually disease-specific.

Determining genetic variances in autoimmune diseases could benefit individuals at high risk as they can be closely monitored. This could also pave the way in finding therapeutic approaches for these diseases. However, despite having this information, developing a treatment or cure is still limited due to non-genetic factors. The interaction between genetic and environmental factors could also explain why there are differences in the clinical manifestations in different patients within the same disease group.

1.2.2 Non-genetic factors in autoimmunity

In twins and familial studies, the low concordance rate of diseases suggests that environmental factors contribute to the aetiology of autoimmune diseases. In the case of SLE, ultraviolet light, drugs and silica exposure have been associated with triggering the disease (D'Cruz, 2000). In coeliac disease, gluten (Troncone and Jabri, 2011) and rotavirus infections (Stene *et al.*, 2006) are suggested to be triggers for the disease. In RA, smoking (Ruiz-Esquide and Sanmarti, 2012), infections (Meron *et al.*, 2010), dietary factors (Hagen *et al.*, 2009) and pollutants (Hart *et al.*, 2009) are important (Tobon *et al.*, 2010). In multiple sclerosis (MS), viral and parasitic infections, smoking and vitamin D have been associated with the disease (Koch *et al.*, 2013) and in Graves' disease, dietary iodine (Laurberg *et al.*, 1991), infections (Marino *et al.*, 2014) and smoking habits (Wiersinga, 2013) play a major role in eliciting the occurrences of the disease. In regards to infections associated with autoimmunity, the microbiome may also impact autoimmunity and has been associated with central nervous system autoimmunity (Ochoa-Reparaz and Kasper, 2014), SLE (Zhang *et al.*, 2014) and T1D (Giongo *et al.*, 2011).

The importance of gene-environment interaction in autoimmune disease is evident, however the precise mechanisms in the aetiology of each disease is still unknown. There seems to be some commonality between the diseases for example, the role of vitamin D and viral infections, however no single environmental factor has been conclusively proven to cause a particular disease.

1.3 Twin studies

Twins studies are invaluable in medical research in providing answers to important questions regarding possible genetic traits on different phenotypes in the aetiology of human disease. MZ and dizygotic (DZ) twin studies assess the variance of phenotypes and attempts to determine how much of this is down to the contribution of genetic or environmental factors. MZ twins share nearly 100 percent of their genetic polymorphisms as they were developed from a single fertilised egg, however significant phenotypic discordance between them may still exist (Fraga *et al.*, 2005). Studying MZ twins is an especially helpful experimental design as any genetic differences would not be an issue between genetically identical twins. The 'normal' twin is a control for the diseased twin, thus minimising the effects of germline genetic variations and some environmental differences as they would have been brought up in very similar environments. A DZ twin, developed from two fertilised eggs, share approximately 50 percent of genes and this level of similarity is also seen in non-twin siblings (Boomsma *et al.*, 2002).

1.4 Type 1 diabetes mellitus

T1D is believed to be a T-cell mediated autoimmune disease that is the result of the interaction of genetic and environmental factors (Atkinson *et al.*, 2014) and can affect an individual at any age (Merger *et al.*, 2013). These factors lead to an altered immune response to destruction of insulin secreting β-cells in the pancreas (Bluestone *et al.*, 2010). This can lead to poor blood glucose control and diabetic patients are predisposed to complications associated with the disease. Hyperglycaemia, hypoglycaemia and diabetic ketoacidosis are examples of acute complications whereas retinopathy, nephropathy, neuropathy and retinopathy are examples of chronic complications (American Diabetes, 2009). Risk factors of developing T1D include increased maternal age at delivery (Bingley *et al.*, 2000), and parental gender, that is the risk to the offspring when the father has T1D is greater than when the other has T1D. Monitoring blood glucose levels is important in order to help prevent hypo- or hyperglycaemia. Currently, treatment is with insulin injections to manage blood glucose levels. Insulin pumps, providing continuous subcutaneous infusions of insulin can also help control blood glucose levels and these pumps reduce the need for multiple insulin injections.

1.4.1 β-cell death

T1D is characterised by the loss of pancreatic β-cell function leading to a deficiency in producing insulin (Nokoff and Rewers, 2013). About 80% of the β-cells are destroyed at the time of clinical symptoms due to insulitis (Notkins and Lernmark, 2001; Cnop *et al.*, 2005). Insulitis is the inflammation of the islets cells in the pancreas containing large numbers of mononuclear cells, $CD8^+$ and $CD4^+$ T cells (Itoh *et al.*, 1993). This infiltration also contains macrophages and B cells, (Willcox *et al.*, 2009) and exposure to soluble mediators released from these cells such as cytokines and nitric oxide (Eizirik and Mandrup-Poulsen, 2001). Natural killer cells are also suggested to be involved in islet inflammation (Dotta *et al.*, 2007). Pancreatic tissue from six T1D patients and 26 healthy controls were analysed. $CD94^+$ natural killer cells were identified in the mononuclear infiltrate of three of the six diabetic patients. Coxsackie B4 was also present in three of the six diabetic patients and the virus extracted from positive islet was able to infect β-cells from healthy donors. These studies imply that both the innate and adaptive immune systems are involved in the proinflammatory process leading to islet cell death.

Although most β-cells are destroyed in T1D patients, residual β-cell mass and insulin secretion has been observed in patients with disease duration of more than 50 years (Keenan *et al.*, 2010). This was assessed by random serum C-peptide levels and response to mixed-meal tolerance test (MMTT) in 411 T1D patients. The C-peptide levels showed that more than 67.4% of the patients had levels in the minimal (0.03–0.2 nmol/l) or sustained range (≥0.2 nmol/l). The authors also reported that from the MMTT, over half of the patients responded by a two-fold or greater rise over the course of the test compared to fasting. Pancreatic islet transplantation is an alternative to replacing β-cells however this procedure is not suitable for all T1D patients and organs are scarce (Correa-Giannella and Raposo do Amaral, 2009).

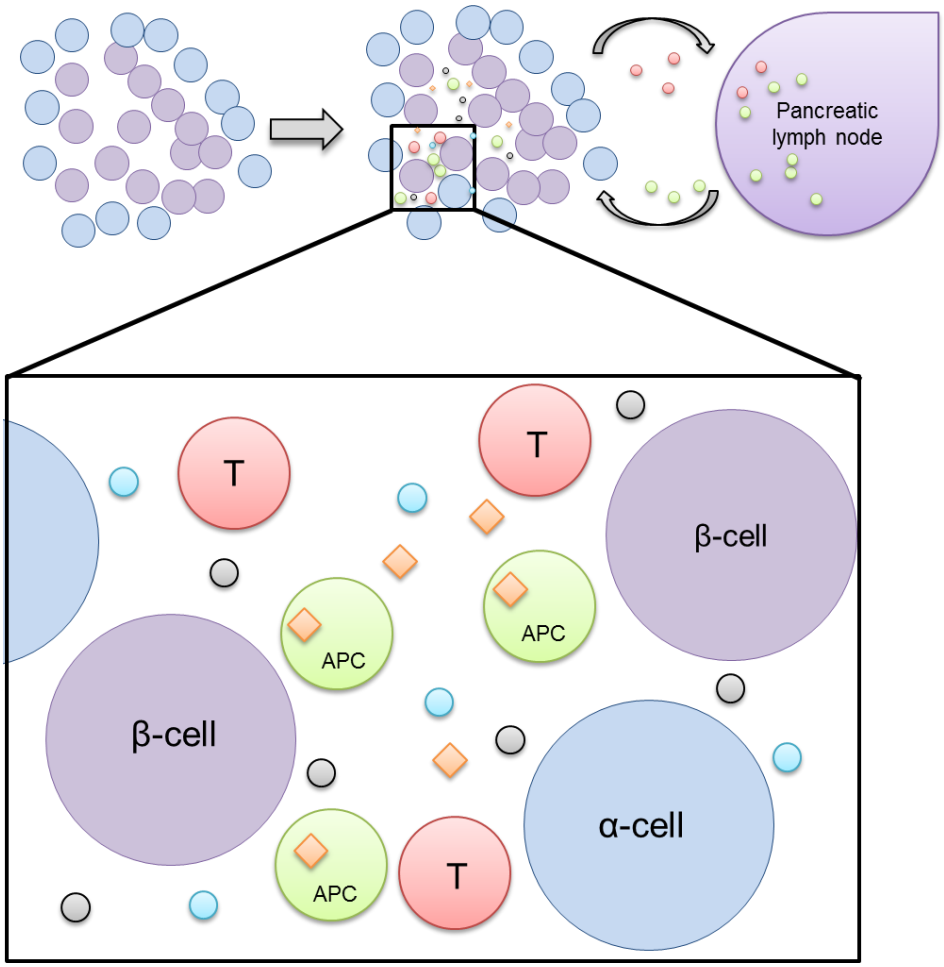


Figure 1.1. Infiltration of immune cells in insulitis. Insulitis is an inflammatory infiltration of immune cells of the islets of Langerhans in T1D. Islet antigens (orange diamonds) are presented to the T cells (red circles) by APCs (green circles) in the pancreatic lymph node. The T cells become activated and attack the β-cells (purple circles) whilst releasing inflammatory factors (small black and blue circles) inducing insulitis. The infiltrate in the pancreatic islets are limited to the β-cells only, leaving the other cells such as the α-cells (blue circles) intact.

1.4.2 Epidemiology of type 1 diabetes mellitus

T1D accounts for 5-10% of all cases of diabetes (American Diabetes, 2009) and in 2000, the worldwide prevalence of T1D was estimated to be 171 million cases (Wild *et al.*, 2004). T1D is seen more commonly in boys and men unlike in other autoimmune diseases where women are more commonly affected (Soltesz *et al.*, 2007). Two decades ago, the incidence of childhood-onset T1D around the world was determined as part of the World Health Organization Multinational Project for Childhood Diabetes (DIAMOND) Project. They reported that the highest incidence rates were found in Northern America and Northern Europe, in particular Finland (36.5/100,000 per year) and Sardinia (36.8/100,000 per year). The lowest incidence rates were found in Asia, Africa and South America, for example in China and Venezuela (0.1/100,000 per year) (Karvonen *et al.*, 2000). There was an overall increase in T1D incidence of 3% per year between 1960 to 1996 in children, with a rapid increase of T1D worldwide in both high and low incidence populations (Onkamo *et al.*, 1999). The incidence of T1D from 23 Epidemiology and Prevention of Diabetes (EURODIAB) centres from 19 countries in Europe, including Germany, Spain and the UK, increased by 3-4% per annum, observed in a 20 year period from 1989 to 2008 (Patterson *et al.*, 2012).

It has also been reported that there was an overall north–south geographical gradient in T1D incidence seen in Europe (EURODIAB, 2000) and in North and South America (Soltesz *et al.*, 2007). There was a seasonality effect in the incidence of T1D with peaks in October to January and troughs in June to August (Moltchanova *et al.*, 2009). However, other studies have shown no evidence of any significant differences in the incidence of T1D during the different seasons, for example in the SEARCH for Diabetes in Youth Study (SEARCH study) in the US (Kahn *et al.*, 2009). It was predicted that the number of T1D cases in children under the age of 15 in Europe will rise between 2005 and 2020 to 160,000 cases (Patterson *et al.*, 2009). The UK has the fifth highest rate of T1D in children according to Diabetes UK, compiling data from the International Diabetes Foundation (Lacobucci, 2013). Compared to Finland and Sweden with incidence rates 57.6/100,000 and 43.1/100,000 respectively, the UK has an incidence rate of 24.5/100,000. The lowest incidence rate is still in Venezuela.

1.4.3 Age-dependent type 1 diabetes incidence

There are different classifications of T1D depending on phenotype at diagnosis. T1D was previously known as juvenile diabetes as many patients were diagnosed at childhood. However, more and more adults are diagnosed with T1D. Diagnosis at childhood suggests a more aggressive role from genetics, as adult-onset diabetes represents slow progress to the disease (Leslie and Delli Castelli, 2004). In cases of adult-onset diabetes, many have previously been diagnosed with type 2 diabetes mellitus (T2D) due to the age at diagnosis, but in combination with autoantibody testing, positivity for autoantibodies against glutamic acid decarboxylase 65 (GAD65), insulinoma antigen 2 (IA-2) and/or zinc transporter 8 (ZnT8) signifies the result of an autoimmune process (Stenstrom *et al.*, 2005). This in turn presents a subset of diabetic patients known as latent autoimmune diabetes of adults (LADA).

LADA is a clinically distinct form of diabetes where the patient presents with an islet autoantibody such as glutamic acid decarboxylase (GADA). They also tend to be younger and leaner (Hawa *et al.*, 2013). LADA patients do not go onto insulin treatment for at least six months post-diagnosis (Leslie *et al.*, 2008), although the time that patients go onto insulin alone, should not solely be used to diagnose LADA (Brophy *et al.*, 2008). There are also differences in the genetic background of T1D and LADA (Vatay *et al.*, 2002). In a cohort of 69 T1D patients and 42 LADA patients, the frequency of the high risk haplotype DR3 was only seen in the T1D group. There was no significant difference between the LADA and control group. The DR3/DQ2 haplotype was also found significantly more often among the T1D patients compared with the control group, whereas there was no difference between the LADA and control group. Immunological differences such as insulin resistance and islet proteins were also identified between T1D and LADA patients (Palmer *et al.*, 2005).

1.5 The role of innate immunity in type 1 diabetes

Immune cell dysfunction in innate immunity has been associated with T1D (Beyan *et al.*, 2003; Grieco *et al.*, 2011). T1D patients have been found to show abnormal monocyte gene expression (Padmos *et al.*, 2008). Using quantitative reverse transcription PCR (qRT-PCR), 25 monocyte activation genes in 30 LADA, 30 juvenile onset T1D, 30 adult-onset T1D and 49 healthy controls, were validated, of which were split into two distinct clusters. One cluster consisted of proinflammatory

genes including *IL1B*, *IL6*, *TNF* and *CXCL2*. The second cluster consisted of genes mainly involved in adhesion, motility, and metabolism such as CCL7, CCL2, and MAPK6. The first cluster was found mainly in the LADA and adult-onset T1D patients, compared to the childhood-onset T1D patients. Whereas the second cluster had the opposite effect and was found more in the childhood-onset T1D patients, indicating that age-dependent T1D incidence have different profiles and immune phenotypes. Of the 25 genes studied, 16 of the genes were then revalidated in 10 MZ twin pairs. Monocyte gene expression abnormality was also seen in identical twins discordant for childhood-onset T1D (Beyan *et al.*, 2010). Aberrant monocyte gene expression was similar in the diabetic and non-diabetic twin compared to normal control twins and healthy control individuals. As there were differences found between the twin pairs discordant for T1D and the normal healthy twins, this suggests that the altered monocyte gene–expression profile could be due to shared gene-environment interaction.

Evidence in support of monocyte gene expression changes in T1D, was reported by Kaizer *et al*. They showed changes in the overexpression of IL1B and MYC in peripheral blood mononuclear cells (PBMCs) (Kaizer *et al.*, 2007). IL1B encodes for IL-1β, a cytokine that recruits inflammatory cells to the islets and have direct cytotoxic effects on β-cells (Sumpter *et al.*, 2011). MYC encodes for a transcription factor which is involved in apoptosis (Laybutt *et al.*, 2002). Perturbed monocyte gene expression in T1D, can distinguish T1D patients at disease onset (Irvine *et al.*, 2012). In this study, peripheral blood monocyte expression was profiled in six healthy subjects and 16 children with T1D diagnosed three months earlier. Using fluorescence-activated cell sorting (FACS) to assess the distribution of different monocyte subsets in the individuals, it was reported that $CD14^{+}CD16^{+}$ monocytes were underexpressed in T1D patients compared with healthy controls. There was also no difference in the proportions of $CD14^{low}/CD16^{+}$ and $CD14^{hi}/CD16^{-}$ populations and the proportion of $CD14^{low}/CD16^{-}$ cells was significantly increased in recent-onset T1D patients.

The innate immune system may also drive the adaptive immune system in T1D (Bradshaw *et al.*, 2009). Monocytes from T1D patients were found to spontaneously secrete IL-1β and IL-6, which induce and expand Th17 cells. The monocytes from T1D patients were also reported to have induced more IL-17-seceting cells from memory T cells than monocytes from healthy controls.

1.6 The role of adaptive immunity in type 1 diabetes

1.6.1 T cells in type 1 diabetes

In the thymus, T cells form the ability to distinguish self from non-self (Kronenberg and Rudensky, 2005) whilst most of the self-reactive cells are eliminated. However, some self-reactive T cells avoid this process and can cause autoimmunity and in the case of T1D, are involved in β-cell destruction (Roep and Peakman, 2011). Dysregulation of T cells has been the main focus of the involvement of the adaptive immunity in T1D (Roep, 2003). In humans, it was reported that defective regulation of T_{regs} was involved in T1D regardless of disease duration (Lawson *et al.*, 2008). Although there was no significant difference in the frequency of $CD4^+CD25^{hi}$ T_{regs} in the T1D patients, the level of suppression by $CD4^+CD25^{hi}$ T_{regs} was reduced in 44 T1D patients compared to 44 controls. An additional finding was that there was no difference in FOXP3 or CD127 expression on $CD4^+CD25^{+hi}$ cells in T1D patients suggesting that an alteration in the balance of T_{regs} and activated T cells does not contribute to the defective suppression in T1D.

The altered frequency of T_{regs} is not associated with T1D but rather altered function of the cells, and has been supported by other studies (Brusko *et al.*, 2005; Lindley *et al.*, 2005). The former study investigated $CD4^+CD25^+$ T cells in 70 T1D patients and 37 healthy individuals. Similar frequencies of $CD4^+CD25^+$ T cells were observed in T1D patients and healthy controls of similar age. There was also no difference between the newly-diagnosed patients and those with long-standing diabetes. The authors also reported defective suppression of the proliferation of effector cells in T1D patients and this was associated with reduced production of IL-2 and interferon-γ. The latter study focussed on $CD4^+CD25^+$ T_{regs} in 21 T1D patients and 15 healthy control individuals. Frequency of $CD4^+CD25^+$ T cells were normal in T1D patients, there was an increase in secretion of interferon-γ and decreased IL-10 production.

Although expression levels of FOXP3 and CD127 were not altered in T1D patients, in one study, messenger RNA (mRNA) levels of T cell genes (CD3G and CTLA4) and B cell genes (CD19 and CD20) were found to be underexpressed in long-term T1D patients, but this change was not detected in new-onset or at-risk T1D groups, suggesting these immune changes occurred during diabetes progression (Han *et al.*, 2012).

1.6.2 B cells in type 1 diabetes

The appearance of T1D-asociated autoantibodies and the involvement of T cells in B cell activation, suggest a role for B cells in the pathogenesis of T1D. There have been reports studying B cell function in T1D (Hinman and Cambier, 2014). A research group reported that B cells can promote T1D by secreting anti-islet autoantibodies to enhance the expansion of islet-reactive CD4 T cells (Silva *et al.*, 2011). However, a study by Wong *et al*. reported that antibody secretion is not essential for some of the diabetes-promoting roles of B cells (Wong *et al.*, 2004). Also, treatment with rituximab can partially preserve β-cell function for one year in T1D patients (Pescovitz *et al.*, 2009). This effect was observed by depleting B cells with the anti-CD20 monoclonal antibody drug and this resulted in reduced destruction of β-cells in patients with T1D of recent onset. The authors also reported improvement in the C-peptide levels and both the glycated haemoglobin level and insulin dose were significantly lowered in the rituximab group in comparison to the placebo group.

Also, an increase in $CD19^+CD5^+$ B cells was reported in T1D children with recent onset of the disease (De Filippo *et al.*, 1997). The authors investigated 16 children with T1D and the presence of $CD19^+CD5^+$ B cells to study the correlation between the number of circulating $CD19^+CD5^+$ B cells and the presence of anti-islet cell autoantibodies. Although the number of $CD19^+CD5^+$ B cells was higher in T1D children compared to the controls, no correlation was found between the B cells and presence of autoantibodies. In contrast, one study reported an increase in $CD20^+CD5^+$ cells in diabetic patients compared to controls but showed a positive correlation between the $CD20^+CD5^+$ B cells subset and GAD autoantibodies (Kadziela *et al.*, 2003). Although there is evidence to suggest that B cells can play a role in T1D, individuals with the absence of B cells can still develop T1D (Martin *et al.*, 2001).

In mice, elevated CD19 expression on B cells promoted presentation of islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP), mediating the expansion of autoreactive T cells specific for antigens to β-cells (Ziegler *et al.*, 2013). The authors also reported that downregulated expression of CD19 significantly diminished the expansion of $CD8^+$ T cells. This suggests that alterations to the CD19 signalling pathway may prevent expansion of islet-invasive T cells and preserve β-cell mass.

1.7 Role of genetic factors in type 1 diabetes mellitus

Genetic influences are important in T1D and this is seen from twins and familial studies. Concordance rates for T1D in identical twins (approximately 50%) are higher than in non-identical twins (Renz *et al.*, 2011; Redondo *et al.*, 2008; Redondo *et al.*, 2001); also the frequency of the disease is higher in siblings of diabetic patients than in the general population (Field, 2002). There have been several large studies set to identify genes involved in T1D, including The Wellcome Trust Case Control Consortium (WTCCC) (Wellcome Trust Case Control, 2007) and The International Type 1 Diabetes genetics Consortium (T1DGC) (Noble *et al.*, 2010). Since then, over 40 distinct genetic loci have been identified to be involved in T1D (Barrett *et al.*, 2009; Pociot *et al.*, 2010). The MHC is the main risk factor in T1D, with the insulin gene and other common variants involved in the aetiology of T1D.

1.7.1 Major histocompatibility complex

T1D is a polygenic disease and is strongly associated with several MHC II genes (Pociot and McDermott, 2002) with the major ones, HLA-DQB1 and HLA-DRB1 on chromosome 6p21.3, accounting for approximately 50% of genetic susceptibility for the disease (Noble *et al.*, 1996). The HLA class II haplotypes DRB1*0401-DQB1*0302 (DR4-DQ3) (Zhou and Jensen, 2013) and DRB1*0301-DQB1*0201 (DR3-DQ2) confer the greatest susceptibility to T1D, and DRB1*1501 (DR15) and DQA1*0102-DQB1*0602 (DQ6.1) provide protection to the disease (Graham *et al.*, 1999). T1D incidence can vary between ethnic groups (Zipris, 2009). For example, in Caucasians, the DR3 and DR4 haplotypes are prevalent, however are rare in Japanese populations (Ikegami *et al.*, 2006). Also, the DR4-DQ (DRB1*0405-DQB1*0401) and DR9 (DRB1*0901-DQB1*0303) susceptible HLA haplotypes were found to be more common in the Japanese but rare in Caucasians (Kawabata *et al.*, 2002). Although the major genetic susceptibility regions have been associated with the MHC class II genes, MHC class I also plays a role in the aetiology of T1D (Nejentsev *et al.*, 2007), particularly HLA-B*39, which was found to contribute to the aetiology of T1D. In this study, samples from 850 T1D affected siblings, 2,048 T1D patients and 1,912 controls were analysed, genotyping 254 polymorphic MHC loci. The strongest associations mapped to the MHC class II genes HLA-DQB1 and HLA-DRB1, but there was also an association around HLA-B. The HLA-A*02 allotype has also been reported to be directly linked to T cell autoreactivity to insulin (Pinkse *et al.*, 2005).

In twin studies, early-onset T1D increases the risk in co-twin of the index twin diagnosed (Hyttinen *et al.*, 2003; Redondo *et al.*, 2001). Also, non-diabetic siblings that have the high risk haplotype shared with their diabetic sibling had increased risk for islet autoimmunity, compared to siblings that did not share the HLA haplotype with their diabetic sibling (Aly *et al.*, 2006). This is useful for identifying a subset of siblings that could be monitored for prevention of islet autoimmunity. In regards to sibling risk for T1D, siblings of patients have a 15-fold higher risk for T1D than in the general population (Steck and Rewers, 2011). HLA-encoded T1D susceptibility only accounts for less than 50% of the inherited disease risk, therefore non-HLA encoded genes such as the insulin gene, *CTLA4* and *IFIH1*, also plays a role in T1D.

1.7.2 Insulin gene

The Insulin-Dependent Diabetes Mellitus 2 (IDDM2) locus, also known as the insulin gene (INS) region, contributes about 10% toward T1D susceptibility (Bennett and Todd, 1996) and is found on chromosome 11p15.5 (Lucassen *et al.*, 1993). This locus has variable number of tandem repeats (VNTR), a type of minisatellite, There are three classes of VNTRs that are divided by the number of repeats. Class I VNTR has an average of 570 bp, class II 1,200 bp and class III 2,200 bp (Bennett *et al.*, 1995). Class I is the shortest of the three and is associated with high risk of developing T1D, whereas class III provides a more dominant protective effect (Julier *et al.*, 1991; Bennett and Todd, 1996).

In a twin study investigating two susceptibility genes in T1D, the insulin gene and HLA-DQB1, it was reported that a particular INS genotype (*Hph I*) was identified in 87.5% of the concordant twins but only 59.5% of the discordant twins (Metcalfe *et al.*, 2001). This suggests that the possession of the high-risk haplotype increases the likelihood of identical twins becoming concordant for T1D.

1.7.3 Non-MHC or insulin loci associated with type 1 diabetes

There are other genetic loci that contribute to disease risk such as the *CTLA4* on chromosome 2q33 (Wang *et al.*, 2014), IL12B on chromosome 5q33-34 (Nisticò *et al.*, 2002), *PTPN22* on chromosome 1p13 (Bottini *et al.*, 2004), the interleukin 2 receptor, alpha (*IL2RA*) (Todd *et al.*, 2007) and interferon induced with helicase C domain 1 (*IFIH1*) genes (Todd *et al.*, 2007).

PTPN22 was shown to regulate type I IFN-induced apoptosis in β-cells (Santin *et al.*, 2011) and a gain-of-function mutation was linked to a decrease in T cell receptor signalling (Vang *et al.*, 2005). A rare variant of the *IL2RA* gene was recently identified to be associated with the disease (Huang *et al.*, 2012). IL2RA encodes CD25, the α-chain of the IL-2 receptor complex. An IL2RA haplotype has been reported to result in diminished IL-2 responsiveness, causing lower levels of FOXP3 expression by T_{regs} (Garg *et al.*, 2012) and a particular single nucleotide polymorphism (SNP) has been shown to be associated with lower circulating levels of CD25, suggesting a predisposition to T1D due to inherited lower immune responsiveness (Lowe *et al.*, 2007). IFIH1, also known as melanoma differentiation-associated protein 5 (MDA5), is a family of intracellular proteins that recognise viral RNA which can mediate the innate immune response (Downes *et al.*, 2010). IFIH1 is part of the interferon regulatory factor 7 (IRF7)-driven inflammatory network (IDIN), which is a network enriched for viral response genes (Heinig *et al.*, 2010). More recently, four rare or low frequency variants within IFIH1 were associated with T1D, indicating the gene is causal (Nejentsev *et al.*, 2009). These variants lowered T1D risk independently of each other in IFIH1, a gene located in a region previously associated with T1D by GWASs. Another SNP of IFIH1 associated with T1D was the IFIH1 rs2111485 genotype (Winkler *et al.*, 2011). Islet autoantibody–positive children with the IFIH1 rs2111485 GG genotype had a faster progression to diabetes than children with the T1D protective GA or AA genotypes. These polymorphisms were associated with the progression to diabetes from autoimmunity but not the development of autoimmunity.

The increasing incidence of T1D and relative stability of HLA gene polymorphisms indicates that this cannot be due to genetic selection pressures and is most likely the result of non-genetic factors.

1.8 Role of non-genetic factors in type 1 diabetes mellitus

Although genetic susceptibility plays a major role in the development of T1D, the rapid rise in T1D incidence cannot be explained by genetics alone. For example, the frequency of the diabetes-risk haplotype HLA-DR3/DR4 was higher in a group of childhood-onset diabetics diagnosed before 1965, compared to the group diagnosed after 1990 (Hermann *et al.*, 2003). Other groups reported similar findings (Gillespie *et al.*, 2004). The frequency of the protective haplotype, HLA-DR15 and HLA-DR1301, was seen to be lower in diabetics diagnosed before 1965 than those

diagnosed after 1990. Therefore, the incidence of T1D is increasing too rapidly to be explained by an increase in genetic susceptibility. This emphasises the influence of environmental factors in the pathogenesis of T1D.

Evidence for non-genetic effects include population studies, in which there is geographical variation in T1D incidence rates. For example, high incidence rates of T1D are found in northern Europe in comparison to Asia and South America (Serrano-Rios *et al.*, 1999). Evidence from migrants studies investigating incidence of T1D in individuals from low incidence countries who moved to high incidence countries, is also suggestive of a non-genetic effect in T1D. One such example is a study investigating children with T1D in Yorkshire (Harron *et al.*, 2011a). A comparison was made between children who originated from the UK and those whose parents were from South Asia, where there is a low incidence rate of T1D. There was a rise in incidence of T1D in both cohorts indicating a role of environmental factors in the development of diabetes. Siblings of diabetic patients have a higher risk of developing T1D and this could also be indicative for non-genetic effects as the siblings would have been raised in a shared environment. This familial clustering (λs) is calculated as the ratio of the risk to siblings over the disease prevalence in the general population (Steck and Rewers, 2011). Non-genetic effects in T1D are also evident by low concordance rates of T1D in MZ twins (Huber *et al.*, 2008). Heritability for T1D was estimated to be 0.88 in the Finnish Twin Cohort with 44 MZ and 183 DZ twin pairs (Hyttinen *et al.*, 2003), indicating an environmental component to T1D.

Triggering mechanisms of the disease could involve the development of autoantibodies and several different hypotheses including hygiene, migration and viral infection which have been suggested to contribute to the aetiology of T1D. It has also been hypothesised that epigenetics could play a role in T1D.

1.8.1 Autoantibodies

The development of the diabetes-associated autoantibodies can be detected years before clinical symptoms occur and in early life (Ziegler and Nepom, 2010). Because MZ twins can be discordant for them they are likely non-genetically determined (Beyan *et al.*, 2012b), indeed, this twin study suggested that both GADA and IA-2A were predominately due to unique environmental events. Antigens associated with T1D are GAD, IA-2, insulin and ZnT8 (Dang *et al.*, 2011; Wenzlau *et al.*, 2007;

Leslie *et al.*, 2001). Autoantibodies against these antigens are detected in patient samples and can differentiate between T1D and T2D in diagnosing a patient with similar phenotypes. Autoantibodies can be predictive of T1D, particularly in adult-onset diabetes (Leslie, 2010). At the time of diagnosis of T1D, 70–80% of patients present with GADA (Notkins and Lernmark, 2001), 65% with IA-2A (Leslie *et al.*, 1999) and 26% with ZnT8A (if negative for the other two) (Wenzlau *et al.*, 2007). Insulin autoantibodies (IAA) are more common in younger children with new-onset T1D than in adults. In diabetic patients, once insulin is administered, detection of IAA is no longer valid as exogenous insulin can prompt insulin antibody responses that cannot be distinguished from autoantibody production (Winter and Schatz, 2011).

Genetic factors play an important role in islet cell autoimmunity. In one study, the siblings of 53 MZ and 30 DZ twin pairs who had T1D were tested for autoantibodies (Redondo *et al.*, 1999). Siblings of MZ twins expressed two or more antibodies more often than siblings of DZ twins. The siblings were also screened for high risk HLA class II haplotypes and it was found that the probability of developing positive autoantibodies was higher among the MZ twin siblings who were positive for HLA DQ8/DQ2 than in those without this genotype. Similar studies that were performed showed that children with high genetic risk also developed islet cell antibodies more often than those at moderate risk (Kupila *et al.*, 2002). Together with HLA genetic predisposition to T1D, positivity for autoantibodies is highly predictive of the disease (Siljander *et al.*, 2009; Winter and Schatz, 2011).

Early seroconversion, from nine months to two years of age, to autoimmunity has been associated with a high rate of progression to T1D (Ziegler *et al.*, 2012; Parikka *et al.*, 2012). This suggests that environmental factors may play a role in the development of T1D very early on. The serum metabolome can also reveal abnormalities in diabetic children at birth (Oresic *et al.*, 2008). The children who developed T1D had reduced serum levels of succinic acid and phosphatidylcholine at birth and increased levels of proinflammatory lysoPCs several months before seroconversion to autoantibody positivity. A later study by the same group also found lower concentrations of methionine in children with early autoimmunity compared to children who developed autoantibodies late and with children who remained autoantibody negative (Pflueger *et al.*, 2011).

1.8.2 Aetiology of type 1 diabetes

The aetiology of T1D is complex as there is evidence of an interplay between genetic and environmental factors (Todd, 2010). Several hypotheses have been investigated which provides evidence for non-genetic factors in the development of T1D.

1.8.2.1 The migration hypothesis

Migration studies are seen to support the role that the environment could account for the rapid increase in the incidence of autoimmune diseases in developed countries. These studies focus on migrants moving from a "low"-incidence area to a "high"-incidence area. For example, children in the UK born to immigrants from Asian countries with a low T1D incidence were found to have an increased incidence of T1D compared to their parents' countries of origin (Bodansky *et al.*, 1992). From this study, data on children from Bradford aged 0-16 years revealed that the incidence of T1D in these children, increased from 3.1/100,000 per year in 1978-81 to 11.7/100000 per year in 1988-90, which was higher than the rates of the children of the indigenous population (10.5/100,000). This study also showed that there was a higher incidence of T1D in Asian males (8.8/100 000 per year) than Asian females (4.9/100000 per year). In a later study, the incidence of the migrants was still rising but not as steeply as reported and the incidence rates did not reach those of the host population (Harron *et al.*, 2011b). Another example of a migration study carried out in the UK was by Raymond *et al*., studying children in Leicestershire (Raymond *et al.*, 2001). Over a 10 year period, the authors reported that T1D incidence rates in 46 South Asian children were similar to those for 263 children who were in the white/other ethnic group.

Children born in Sweden have an increased risk for T1D where their origin is from low incidence countries such as East Asia, Eastern Europe and Latin America (Soderstrom *et al.*, 2012). In a Swedish study cohort of children aged between 6-25 years were split into four categories: international adoptees, immigrants, Swedish-born with foreign-born parents and a comparison group with Swedish-born parents. The odds ratios for T1D were lower in the groups with an origin in low incidence regions compared to the group with Swedish-born parents. Further evidence of exposure to environmental factors in T1D was from a study investigating the age at onset of T1D in individuals aged between 0 and 30 years of age from the host population of Sweden compared to those with foreign born parents (Hussen *et al.*,

2013). The findings showed there was an almost identical pattern with a shift towards lower age at onset in both cohorts, pointing to increased exposures in early life that could initiate or accelerate β-cell destruction. Conversely, data collected on Japanese children diagnosed with T1D in Hawaii, showed incidence rates were comparable with rates in Japan (Patrick *et al.*, 1997). In this study, 113 children of different ethnicity diagnosed with T1D were identified. The incidence rates between the different groups varied greatly. Children who were part Hawaiian had the highest incidence rates (15.34-16.58/100,000), followed by Caucasian children (6.21-6.71/100,000) then Filipino children (3.66-3.96/100,000) and Japanese children (2.85-3.08/100,000). The low incidence rates in the Filipino and Japanese children in Hawaii suggest genetics may play a larger role in developing T1D than environmental factors.

1.8.2.2 The hygiene hypothesis

It has been observed that the incidence of T1D is higher in more industrialised countries with higher standards of living, and therefore higher levels of hygiene, than those with lower standards of living. This has led to the "hygiene hypothesis." (Bach and Chatenoud, 2012). This hypothesis suggests that one of the causes of T1D is due to reduced exposure to infections, resulting in less protection from infectious agents (Bach, 2002). Maternal antibodies against enteroviruses were measured and it was found that increased frequency of enteroviral antibodies correlated with low risk of T1D development (Larsson *et al.*, 2013). Enteroviruses may have immunoregulatory effects in humans as they have been found to be associated with low risk of IgE-mediated allergic sensitisation (Seiskari *et al.*, 2007). The role of the microbiome has recently been claimed to influence T1D development (Dunne *et al.*, 2014) as the gut microbiota can modulate the function of the immune system (Vaarala, 2013). This may explain why there is an increase in incidence of T1D in the more developed countries. This hypothesis has been supported by studies carried out in farms focusing on children brought up in this environment (Heikkinen *et al.*, 2013) where a variety of non-pathogenic microorganisms may have a protective effect on T1D. However, it has been suggested that it is the depletion of a certain species from the ecosystem of the human body, that may lead to allergic disease, contradicting suggestions that immune disease were the results of infections (Parker, 2014).

1.8.2.3 The diet hypothesis

Dietary factors can play an important role in the development of T1D. For instance, breastfeeding has been suggested to play a protective role and that the introduction of cow's milk has an influence on the risk of developing diabetes (Pereira *et al.*, 2014; Hummel *et al.*, 2014; Lund-Blix *et al.*, 2014). The early introduction of cow's milk to an infant introduces antigenic bovine insulin which has been shown to increase an immune reaction by increased antigen presentation and increased β-cell stress (Virtanen *et al.*, 2000), which later on may turn into an autoimmune reaction against own insulin secretion (Vaarala *et al.*, 2002). The early introduction of other components in cow's milk might also play a role and cause immune reactions in genetically predisposed individuals (Knip *et al.*, 2010) in the same manner such as early introduction of gluten-containing food is a potential risk factor for developing T1D (Ziegler *et al.*, 2003). While gluten-containing foods are a major risk factor for the development of coeliac disease, an autoimmune disease genetically closely related to T1D, the current evidence does not favour a role for milk and autoantibody risk (Knip *et al.*, 2014). Knip *et al*. performed a double-blind randomised clinical trial of 2159 infants with HLA disease susceptibility and a first-degree relative with T1D. The infants were either weaned to a hydrolysed casein formula or to a conventional cows' milk-based formula. The absolute risk of positivity for two or more islet autoantibodies was 13.4% among those randomized to the casein hydrolysate formula (n = 139) compared to 11.4% among those randomized to the conventional formula (n = 117). The authors had found that the use of the hydrolysed formula, did not reduce the incidence of diabetes-associated autoantibodies after seven years.

In addition, children who were breast-fed for significantly shorter periods of time developed T1D (Kimpimaki *et al.*, 2001a). Kimpimaki *et al.* studied children with the HLA susceptibility allele HLA-DQB1*02/*0302 and monitored them for islet cell antibodies. Children who had been breast-fed for at least four months had lower risk of seroconversion to positivity for IA-2A compared to infants who had been breast-fed for less than two months. These findings suggest that short-term breastfeeding predispose young children who are genetically susceptible to T1D. Evidence from other studies suggests that prolonged breastfeeding can offer protection against development of T1D (Pereira *et al.*, 2014; Alves *et al.*, 2012). This could be due to constituents of cow's milk or breastfeeding confers protection by the transfer of immunity from mother to child.

1.8.2.4 Contribution of vitamin D

There is a seasonality effect in the diagnosis of T1D in that it is more common in the winter months. This may be due to sunshine and vitamin D doses (Hitman *et al.*, 1998). It has also been shown that there is a north-south geographical gradient in the incidence of T1D, implying that there is an inverse correlation between the amount of sunshine and T1D incidence (Zhou *et al.*, 2013). In T1D patients, it has been reported that at diagnosis there are lower serum concentrations of the active form of vitamin D, 1,25-dihydroxyvitamin D (1,25(OH)2D) and its precursor 25-hydroxyvitaminD (25(OH)D), compared to normal controls (Cooper *et al.*, 2011). The group measured 25(OH)D concentrations in 720 cases and 2610 control plasma samples and tested genetic variants influencing 25(OH)D metabolism. They had found three key 25(OH)D metabolism genes, CYP27B1, DHCR7 and CYP2R1, that showed consistent evidence of association with T1D risk, indicating a genetic etiological role for vitamin D deficiency in T1D. These findings were supported by a study which investigated variants in CYP27B1 (Bailey *et al.*, 2007). Conversely, in one study a group followed a cohort of children at increased risk of diabetes and recorded vitamin intake and obtain measurements in plasma 25(OH)D levels (Simpson *et al.*, 2011). They had discovered that neither vitamin D intake nor 25(OH)D levels were associated with the risk of islet autoimmunity or progression to T1D.

There has been conflicting evidence to suggest that vitamin D levels during pregnancy have an effect on the risk of developing T1D. Vitamin D supplementation in particular, was associated with a decreased frequency of T1D (Hyppönen *et al.*, 2001). Taking vitamin D supplements during pregnancy can also lower the risk of the child developing T1D (Sørensen *et al.*, 2012). The group compared serum concentrations of 25(OH)D in 109 women whose child went on to develop T1D and 219 healthy control women. The odds of T1D was more than two-fold higher for the offspring of women with low levels of 25(OH)D compared to the control cohort. However, contrasting results showed that there was no difference in concentrations of 25(OH)D in serum samples from mothers whose children went on to develop T1D (Miettinen *et al.*, 2012).

1.8.2.5 The virus hypothesis

As mentioned earlier, there is seasonal variation in the diagnosis of diabetes, with peaks in the autumn and winter months. High risk infants tested positive for

autoantibodies in the colder months (Kimpimaki *et al.*, 2001b) suggesting that infectious agents may play a role in the pathogenesis of T1D (Craig *et al.*, 2013). Viruses have been thought to trigger T1D, notably rubella and enteroviruses (Hyöty and Taylor, 2002; Yeung *et al.*, 2011) in which these infections can occur several years before the onset of clinical diabetes (Lönnrot *et al.*, 2000). One particular enterovirus, Coxsackie B4, has been linked to T1D. A study found that 64% of children who were diagnosed with T1D under the age of six, were positive for enterovirus RNA in serum samples (Clements *et al.*, 1995). Enterovirus RNA has also been detected in sera of prediabetic children (Nairn *et al.*, 1999), pancreas (Dotta *et al.*, 2007) and small intestine (Oikarinen *et al.*, 2008). The rate of progression from islet autoimmunity to overt T1D was increased after the detection of enteroviral RNA in serum samples (Stene *et al.*, 2010).

Genetic predisposition to T1D is also associated with enteroviral antibodies. It has been reported that children with the high risk haplotype HLA-DR3/DR4, had higher Coxsackie B4 antibody levels than children carrying the protective HLA-DR2 allele (Sadeharju *et al.*, 2003). Furthermore, enteroviruses that were detected in the blood of newly-diagnosed diabetic patients were also found in their siblings and parents (Salvatoni *et al.*, 2013). Besides genetic predisposition, the appearance of autoantibodies was also related to enterovirus positivity. Enteroviral RNA was more likely to be detected in T1D than in control subjects before the appearance of autoantibodies (Oikarinen *et al.*, 2011) and this risk effect was seen to be stronger in boys than in girls. Conversely, there is evidence that has shown that there were no correlations found between the presence of enteroviruses and the development of autoantibodies (Simonen-Tikka *et al.*, 2011; Cinek *et al.*, 2014).

There is lack of evidence to support one specific environmental trigger for T1D, even though there are numerous reports on proposed triggers such as early infant diet (Pereira *et al.*, 2014), viral infections (Craig *et al.*, 2013) and vitamin D supplements (Cooper *et al.*, 2011). Autoantibodies are non-genetically determined and highly predictive of the disease. However, while these autoantibodies can reflect the presence of the disease process, there is no evidence to conclude that they cause the disease. Non-genetic events could also affect gene expression through epigenetics.

1.8.3 Epigenetics

Epigenetics is defined as mitotically heritable changes in gene expression that do not directly alter the DNA sequence (Bird, 2007; Irizarry *et al.*, 2009). As regulators of transcription, epigenetic mechanisms play a necessary role in maintaining normal growth, development, differentiation and genome stability (Bell and Spector, 2011). Genetics play a major role in diseases, however the environment can modify this introducing a complex interaction between the two (Figure 1.2). Environmental factors can also affect DNA by modifying epigenetic factors (Liu *et al.*, 2008). For example in mice, paternal diet has been shown to affect cholesterol and lipid metabolism in the offspring (Carone *et al.*, 2010).

The best-characterised epigenetic modifications or marks are DNA methylation (Figure 1.3) and histone post-translational modifications (Figure 1.4) (Petronis, 2010). Epigenetic dysregulation has been associated with several human diseases, most notably cancer (Laird, 2003; Feinberg, 2007). Studying disease in twin pairs is particularly useful in the field of epigenetics. The discordance seen between MZ twins may be determined by epigenetic factors operating on genetic expression (Poulsen *et al.*, 2007). Studying identical twins in epigenetics is crucial in removing genetic confounding factors and helps to determine the extent of epigenetic heritability and stability (Bell and Saffery, 2012; Petronis, 2006). Differences have been found in DNA methylation and histone acetylation between young and elderly MZ twins (Fraga *et al.*, 2005). The younger pairs were indistinguishable in their epigenetic markings and the older MZ pairs had substantial variations. This then suggests an explanation for age-accumulating epigenetic modifications. These differences contribute to the non-genetic influence involved in T1D pathogenesis as in DZ twins, there are differences not only in the DNA sequence, but also because they originated from epigenomically different zygotes (Kaminsky *et al.*, 2009).

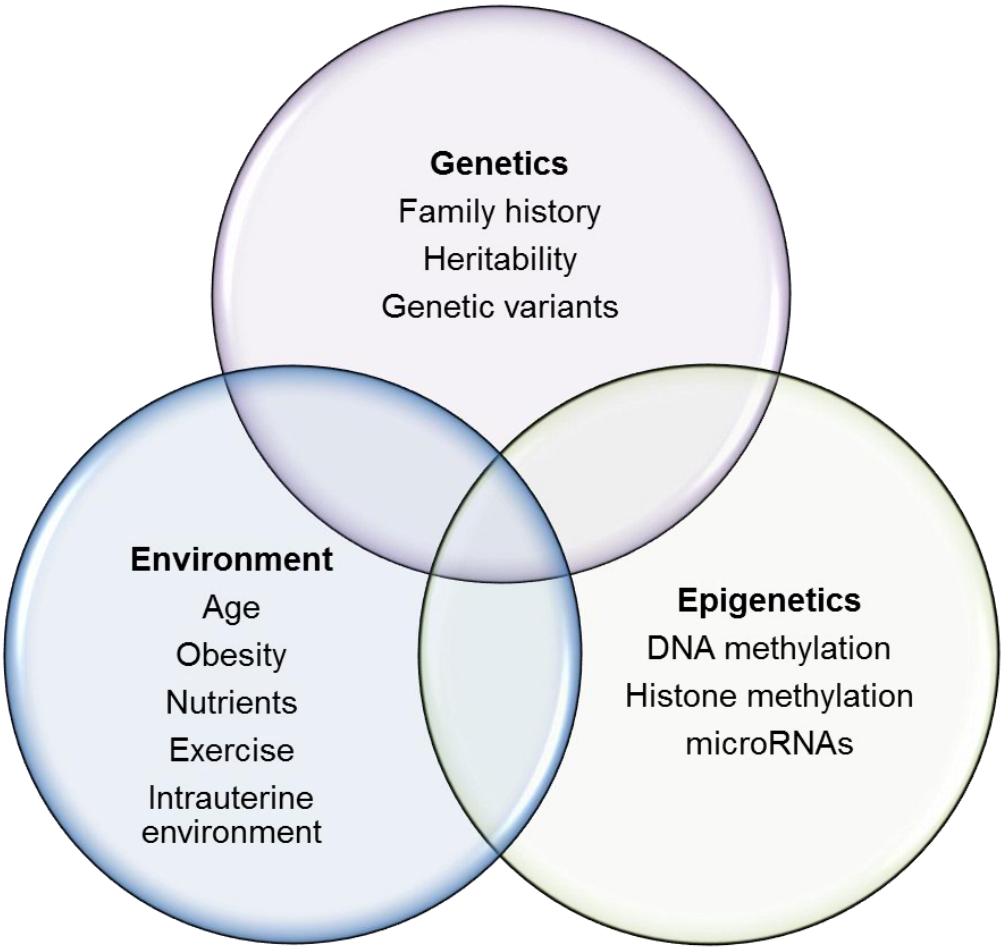


Figure 1.2. Interaction between genes, environment and epigenetics in disease. The genome can give rise to many phenotypes. Although genetics, epigenetics and the environment can affect phenotype outcomes independently, it is the complex interaction that gives rise to diseases such as T1D. Evidence for this includes MZ twin studies in which disease concordance was not 100%. Factors such as age and dietary nutrients have been shown to affect the epigenome and some of these epigenetic changes can occur *in utero*.

1.8.3.1 Epigenome-wide association studies

As DNA methylation is so vital in normal development and irregularities are seen in disease, investigating these epigenetic marks requires high-quality and reliable approaches. There are several different profiling technologies for epigenome-wide association studies (EWAS) (Table 1.1) (Michels *et al.*, 2013). Also the emergence of next-generation sequencing technologies has allowed researchers to obtain information of genome-wide variation at high resolution (Bock *et al.*, 2010). Here they are able to assess epigenetic modifications at multiple pre-determined targets across the genome such as DNA methylation and post-translational histone modifications (Beck and Rakyan, 2008).

The choice of tissue and cohorts is important in EWAS design as disease-associated epigenetic variation can be tissue-specific (Lowe *et al.*, 2013) and variation in methylation will be different in a cohort of MZ twin pairs compared to diseased singletons and their age-matched controls (Rakyan *et al.*, 2011b). Several resources for human genome-wide tissue-specific DNA methylation profiles include methylation profile of DNA (mPOD) (Rakyan *et al.*, 2008), where the authors had generated reference human genome-wide DNA methylation profiles for 13 normal somatic tissues, placenta and sperm. There are also published reference profiles of human chromosomes 6, 20 and 22 (Eckhardt *et al.*, 2006).

As our knowledge in the field increases, this is all applied to developing smaller, faster and cheaper technologies to investigate DNA methylation and histone modifications. The integration of EWAS and GWAS data can assist in identifying genomic and biological pathways that are associated with disease (Ke *et al.*, 2013).

Table 1.1. Different epigenetic profiling technologies. Methylated DNA immunoprecipitation sequencing (MeDIP-seq); chromatin Immunoprecipitation Sequencing (ChIP-seq); Illumina Infinium HumanMethylationBeadChip450K (Illumina450K); bisulfite sequencing (BS-seq); reduced representation bisulphite sequencing (RRBS).

Method	Description	Reference
MeDIP-seq	MeDIP-seq involves immunoprecipitating DNA containing methylated cytosines using a monoclonal antibody raised against 5-methylcytosine (5-mC). The purified fraction of methylated DNA then undergoes next generation sequencing.	(Taiwo *et al.*, 2012)
ChIP-seq	A protein of interest is enriched by immunoprecipitation from cross-linked cells, along with its associated DNA. Enriched genomic DNA sites then undergoes sequencing.	(Landt *et al.*, 2012)
Illumina450K	Microarray that interrogates DNA at over 485,000 CpG sites that includes 96% of CpG islands, 99% RefSeq genes and non-CpG methylated sites. Requires bisulfite converted DNA and is the successor of the Illumina27K BeadChip, which measures DNA methylation in over 27,000 CpG sites.	(Sandoval *et al.*, 2011)
BS-seq	BS-seq involves whole-genome sequencing of CpG sites. Requires bisulfite converted DNA which is fragmented and amplified. Currently considered to be the 'gold standard' technique for the detection of 5-mC. Provides highest level of coverage and resolution.	(Li and Tollefsbol, 2011)
RRBS	Reduced representation bisulphite sequencing involving single base resolution methylation analysis using bisulphite DNA sequencing of a representative part of a genome.	(Meissner *et al.*, 2005)
MBD affinity purification	DNA containing methylated CpG is immunoprecipitated using an MBD column.	(Wielscher *et al.*, 2011)

1.8.3.2 DNA methylation

DNA methylation is essential for regulating the expression of mammalian genes. It is the covalent addition of a methyl group to the 5' carbon in a cytosine-phosphate-guanine (CpG) dinucleotide from the methyl donor S-adenosylmethionine (Laird, 2003). DNA methylation is associated with gene silencing and has been reported to be essential for embryonic development (Reik *et al.*, 2001), genomic imprinting (Reik and Walter, 2001; Heijmans *et al.*, 2008) and X-inactivation in mammals (Yang *et al.*, 2011; Augui *et al.*, 2011; Sharp *et al.*, 2011). Environmental agents that are associated with altered DNA methylation include dietary sources such as folic acid, methionine and choline (Poirier, 2002; Cooney *et al.*, 2002) and smoking (Besingi and Johansson, 2014).

There are regions with higher CpG density called CpG islands (CGIs). CGIs are defined as the regions of DNA of at least 200 bp, with CG content of greater than 50% and are mainly unmethylated (Gardiner-Garden and Frommer, 1987). However, different algorithms have been developed to identify such regions (Bock *et al.*, 2006; Zhao and Han, 2009; Han and Zhao, 2009). CpG sites located at the non-CGIs are mostly methylated in mammals (60-80%) (Smith and Meissner, 2013). CGIs can be found in the promoter regions of genes, and CpG methylation of these gene promoters is associated with transcriptional silencing (Klose and Bird, 2006). Approximately half of the CGIs are not restricted to the promoter region of genes and these CGIs are referred to as 'orphan islands' (Illingworth *et al.*, 2010). It is now shown that DNA methylation is not restricted to the CGIs. They can be identified in regions adjacent to the CGIs also known as CGI shores (Irizarry *et al.*, 2009) and these are usually found approximately 2 kb up- or down-stream (Irizarry *et al.*, 2009). Although cytosine is typically methylated at the C5 position on the carbon, it can also be converted to 5-hydroxymethylcytosine (5-hmC), resulting from the enzymatic conversion of 5-mC into 5-hmC by the ten-eleven translocation (TET) family of cytidine oxygenases (Branco *et al.*, 2011). DNA methylation can also occur in non-CpG sites (Ramsahoye *et al.*, 2000). Ramsahoye *et al*. found that methylation of cytosines can occur at CpA and even CpT sites in embryonic stem cells.

Although every cell has the same genome, the body consists of a large number of differentiated cells and tissue-types. DNA methylation shows tissue specificity (Rakyan *et al.*, 2008), inter-individual variation and a bimodal distribution within the

MHC (Rakyan *et al.*, 2004). DNA methylation patterns vary and some examples are ageing-associated differentially methylated regions (aDMRs) (Rakyan *et al.*, 2010), tissue-specific-DMRs (tDMRs) (Rakyan *et al.*, 2008), methylation variable positions (MVPs) and differentially-methylated positions (DMPs) (Rakyan *et al.*, 2011b).

Immune cell development has many steps that are regulated by a network of transcriptional factors and epigenetic machinery (Kanno *et al.*, 2012; Zilbauer *et al.*, 2013). Current knowledge on the role of DNA methylation on gene expression and its effect on normal cell development is limited. In T cells, deletion of DNMT1 in early double-negative thymocytes led to impaired survival of TCRαβ^{+} cells and the generation of atypical CD8+TCRγδ^{+} cells in mice (Lee *et al.*, 2001). This study suggests that DNMT1 is crucial in normal development of T cells. DNA methylation and gene expression changes were also studied in B cell development in healthy individuals (Lee *et al.*, 2012b). DNA methylation changes were associated with effects on gene expression during early lineage commitment mainly in the gene body rather than the promoter sites, for example at CGIs and transcription factors such as EBF1 and PAX5 which are both regulators in B cell development. As the multipotent progenitors committed to pre-B cells, 79% of the 2966 DMRs were demethylated. The effect of DNA methylation on gene regulation was less in the later stages of B cell development. Methylation changes reported to be observed outside the promoter sites suggests that these regions play a major role in regulating gene expression.

1.8.3.3 DNA methyltransferases and methyl-CpG-binding proteins

DNA methylation is mediated by a group of DNA methyltransferases (DNMTs): DNMT1, DNMT3a, DNMT3b, DNMT3L and DNMT2 (Ramsahoye *et al.*, 2000). DNMT1 is responsible for the maintenance of methylation patterns during replication and for recognising and converting the unmethylated strand from a double stranded hemimethylated CpG pattern into a fully methylated pattern. DNMT3a and DNMT3b are responsible for de novo methylation (Bird, 2002). DNMT3L forms a heterotetramer with DNMT3a and stimulates the activity of the de novo DNA methylation mechanism (Jia *et al.*, 2007). The role DNMT2 plays in DNA methylation is still unknown, however, it has been shown to have a role in the cellular response to stress (Thiagarajan, 2011).

Along with DNMTs, methyl CpG binding domain (MBD) proteins also regulate methylation (Jaenisch and Bird, 2003). The MBD protein family (MBD1, MBD2, MBD3, MBD4 and MECP2) has a common motif that binds to cytosines that are methylated. They then recruit chromatin inactivation complexes, which leads to transcriptional silencing. DNA demethylation can occur as either passive or active DNA demethylation, as seen in the epigenetic reprogramming of germ cells (Reik *et al.*, 2001).

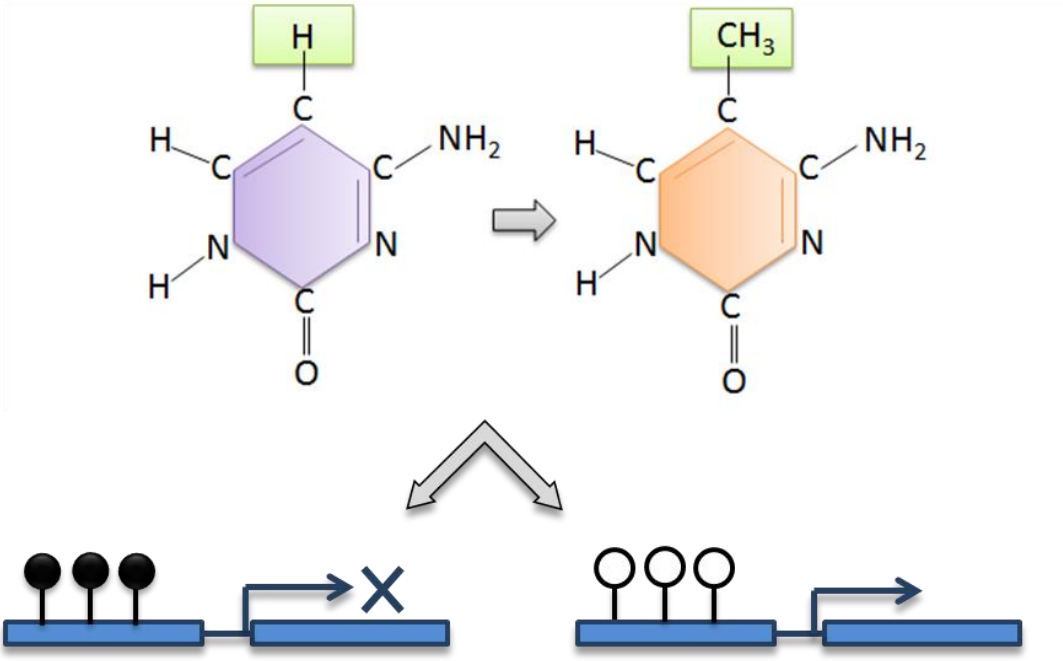


Figure 1.3. DNA methylation and gene expression. A simplified schematic of DNA methylation and its effect on gene expression. DNA methylation occurs by the covalent addition of a methyl group to the 5' carbon of a cytosine nucleotide. Methylated CpG sites (filled lollipop) are associated with gene silencing, whereas unmethylated sites (unfilled lollipop) are associated with transcriptional activity. In the case of hydroxymethylation, at the 5' carbon position, the hydrogen molecule is replaced by a hydroxymethyl group.

1.8.3.4 DNA methylation in type 1 diabetes

Little is known about epigenetic mechanisms in the aetiology of T1D. Recently, evidence is emerging that epigenetic regulation can play a role in triggering T1D. Rakyan *et al*. performed an EWAS in $CD14^+$ monocytes from T1D-discordant MZ twin pairs (Rakyan *et al.*, 2011a). They identified 132 different CpG sites that were significantly linked with diabetic condition. Some of the genes they found to be hypomethylated or hypermethylated such as *GAD2* and *HLA-DQB1,* are known to be associated with T1D. In the same study, T1D–associated methylation variable positions (T1D-MVPs) were found in islet autoantibody positive individuals many years before clinical diagnosis, indicating that they arise early in the disease process. Investigating monocytes is useful as they have been shown to display abnormal gene expression in T1D patients (Padmos *et al.*, 2008). Another twin study had reported 88 CpG sites that showed significant methylation changes in MZ pairs discordant for T1D (Stefan *et al.*, 2013). The group profiled Epstein-Barr virus (EBV) immortalized B cell lines from three MZ twin pairs discordant for T1D and six MZ twin pairs concordant for the disease using the Illumina Infinium HumanMethylation27K BeadChip. Functional analysis of distinct CpG methylation profiles in the T1D samples, showed differential DNA methylation of immune response and defence response pathways between the diabetic and non-diabetic twins. As with the study carried out by Rakyan *et al.*, the data mapped several known T1D associated genes, such as HLA, INS and IL-2RB.

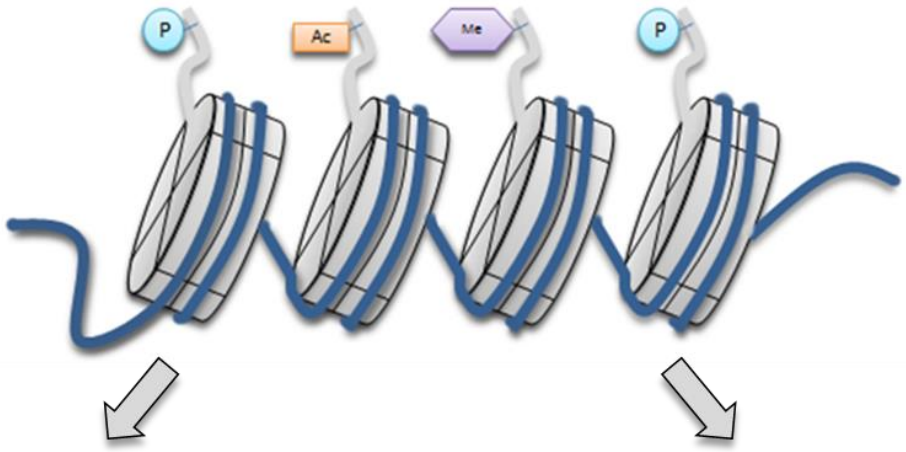
Methylation patterns in the insulin promoter has been described by Fradin *et al.* (Fradin *et al.*, 2012). They had found differences in methylation between T1D patients and non-diabetic controls and identified a 3-CpG-hypomethylation pattern that seemed to be present only in T1D patients. These three CpG sites are proximal to the transcription start site in the insulin promoter gene and therefore could be predictive of T1D. Moreover, 19 prospective CpG sites were identified that correlated with the time of onset of nephropathy, a major complication of T1D (Bell *et al.*, 2010). Out of the 19 CpG sites, one CpG site was found to be hypermethylated. This mark was in the *UNC13B* gene which is associated with the risk of diabetic nephropathy.

1.8.3.5 Histone post-translational modification

Alterations in the chromatin structure affect the expression and repression of genes by enzymatic modification of core histones (Li, 2002). The modifications of the histones result in conformational changes of the chromatin that alter the access of promoters for transcription factors. The nucleosome, which is the basic repeat unit of chromatin, consists of two copies each of histones H2A, H2B, H3, and H4, wrapped by 146 base pairs of DNA (Li, 2002). The nucleosomes are then arranged to form chromatin fibres. Epigenetic modifications in the chromatin structure consist of lysine, arginine and serine residues in the N-terminal tails of core histones being post-translationally modified by acetylation, methylation, ubiquitination or phosphorylation (Strahl and Allis, 2000). These modifications alter the interaction between the histones, DNA and nuclear proteins, therefore affecting gene transcription (Talbert and Henikoff, 2006).

The best characterised histone modification is lysine acetylation. N-acetylation prevents the ε-amino group from binding to DNA, helping to unwind the compact chromatin, while deacetylation of the terminal lysine residues contribute**s** to gene silencing (Bernstein *et al.*, 2007). Acetylation is regulated by histone acetyltransferases (HATs), which add an acetyl group to the histone tails, and histone deacetylases (HDACs) which remove the acetyl group (Peserico and Simone, 2011). There are 18 HDACs in humans and they require the interaction of MBDs for histone deacetylation (Marks, 2009). Examples of gene activation through these modifications include acetylation of histone H3 at lysines 9, 14, 18, 23, and 56. Acetylation of histone H4 at lysines 5, 8, 12 and 16 is also associated with transcriptional activation (Biel *et al.*, 2005). It was thought that DNA methylation and histone modification are interdependent processes. However, MBDs and HDACs can form complexes to induce gene silencing. Another main histone modification is histone methylation. Histone methylation affects lysine or arginine residues in the tails and is regulated by histone methyltransferases, lysine-specific demethylase 1 and Jumonji C domain-containing histone demethylase (Marks, 2009). Methylation of histone H3 at lysines 4, 36 and 79 is associated with enhanced gene expression, whereas methylation at lysine 9, 27 and H4 at lysine 20 is associated with gene silencing (Brooks *et al.*, 2010; Hon *et al.*, 2009). Histone phosphorylation occurs at H3 ser10, ser28 and of histone H4, serine 1 (Biel *et al.*, 2005).

Histone modifications have also been studied in normal immune cell development. In B cells, during the transition of pre-pro-B cell to pro-B cells, E2A-associated genes become monomethylated at lysine 4 on H3 (H3K4me) (Lin *et al.*, 2010). Histone modifications are also involved in monocyte to macrophage differentiation (Tserel *et al.*, 2010). During differentiation, monocyte-specific CD14, CCR2 and CX3CR1 had decreased levels of H3K4me3 but increased levels on dendritic cell specific TM7SF4/DC-STAMP, TREM2 and CD209/DC-SIGN genes. The FoxP3 locus has been reported to be associated with varying histone methylation in T_{regs} (Floess *et al.*, 2007). Acetylation at H3 and H4 and trimethylation at H3 were associated with the conserved FoxP3 region in $CD25^{+}CD4^{+}$ T_{regs} but not in $CD25^{-}$ $CD4^{+}$ T cells. Also the authors reported that in $CD25^{-}CD4^{+}$ T cells, the FoxP3 locus was packed in a more condensed, inaccessible chromatin structure in contrast to $CD25^{+}CD4^{+}$ T_{regs}, where the locus was located within open chromatin which supports evidence that FoxP3 is expressed in T_{regs}. The study concluded that the stabilised FoxP3 expression is needed for normal development of a permanent suppressor cell lineage in T cells.



Transcriptional repression

H3K9me2
H3K9me3
H3K27me2
H3K27me3
H4K20me3

Transcriptional activation

H3S10phos
H3K4me2
H3K4me3
H3K9ac
H3K14ac
H3K27me1
H3K27ac
H3K79me2
H3K36me3
H4K20me1
H4K5ac
H4K12ac

Figure 1.4. Chromatin structure with histone modifications. Simplified schematic of the chromatin structure with histone modifications. Different modifications result in conformational changes to the chromatin. The nucleosome is made up of two copies of each histones H2A, H2B, H3 and H4, wrapped by 146 base pairs of DNA. Methylation of H3 on lysine at position 4, 36 and 79 leads to an actively transcribed open chromatin structure. Methylation at H3 on lysine at position 9 and 27 leads to transcriptional repression. P: phosphorylation, Ac: acetylation, Me: methylation.

1.8.3.6 Histone modification in type 1 diabetes

Miao *et al*. used ChIP-chip to compare genome-wide histone H3K9me2 patterns in peripheral lymphocytes and monocytes from T1D patients and normal subjects (Miao *et al.*, 2008). In lymphocytes, they found a significant increase in H3K9me2 in *CTLA4*, a high risk gene for T1D. It was suggested that the increase in the H3K9me2 mark in the promoter site shows a recent history of transcriptional activation and therefore, its mark would result in suppression of subsequent transcription due to its repressive role. The same group also investigated post-translational modifications at the promoter and enhancer regions of T1D susceptible genes in lymphocytes and monocytes (Miao *et al.*, 2012). They reported variations in H3K9Ac levels at the upstream regions of HLA-DRB1 and HLA-DQB1 in T1D monocytes compared to controls. H3K9Ac is a mark of actively transcribed promoters and HLA-DRB1 and HLA-DQB1 are highly associated with T1D. Hence, these results suggest that the H3K9Ac status of HLA-DRB1 and HLA-DQB1 show an important role of the chromatin status through genetic and epigenetic architecture to assess the functional association to T1D susceptibility.

T cells are involved in the destruction of β-cells therefore an assumption would be that epigenetic marks could alter the function of these cells. In one particular study, HDAC expression was found to be downregulated in $CD4^+$ T cells in T1D (Orban *et al.*, 2007). This suggests that the $CD4^+$ T cells were hypo-responsive in T1D patients. Hyperglycaemia was also observed to affect histone methylation (Brasacchio *et al.*, 2009). The upregulation of the NF-κB-p65 gene, due to the histone methylation of the promoter region of the gene, was a result of prior hyperglycaemia. The effects that epigenetics could have on metabolic memory and hyperglycaemia was further reviewed by Keating *et al* (Keating and El-Osta, 2012). In a twin study, gene expression differences in skin fibroblasts were reported in MZ twin pairs discordant for T1D (Caramori *et al.*, 2012). The group had found differences in the expression levels of genes regulating epigenetic processes. Histone lysine methyltransferase (SET 7), H3K4 methyltransferase and HDAC 8 were downregulated, whereas HDAC 4 was upregulated in the T1D twin. In a mice study, curcumin treatment was shown to increase acetylation of histone H3 (Tikoo *et al.*, 2008). Curcumin has been used to treat other disease such as cancer and offers protection against nephropathy, a serious T1D complication.

1.8.3.7 Non-coding RNA mediated gene-silencing

Gene expression can also be regulated by non-coding RNAs (ncRNAs). ncRNAs include short microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs) and large intergenic non-coding RNAs (lincRNAs) (Zaratiegui *et al.*, 2007). miRNAs are genome-encoded 21 to 23 bps RNAs involved in post-translational regulation of gene expression (Tomankova *et al.*, 2011). miRNAs target the RNA-induced silencing complex sites found at the 3' untranslated region of specific mRNAs. This leads to degradation or translational repression and affects the regulation of cellular processes such as embryonic development, cell differentiation, cell cycle and apoptosis (El Gazzar and McCall, 2011). More than 700 miRNAs have been identified in mammalian cells, and aberrant miRNA expression can lead to autoimmune diseases (Ha, 2011). miRNAs can also regulate DNA methylation and histone modifications (Tomankova *et al.*, 2011).

1.8.3.8 microRNA regulation in type 1 diabetes

There is evidence that miRNAs are involved in the pathogenesis of T1D (Guay *et al.*, 2011; McClelland and Kantharidis, 2014; Kumar *et al.*, 2012; Mao *et al.*, 2013). Sebastiani *et al*. analysed expression levels of miR-326 in peripheral lymphocytes from T1D patients possessing autoantibodies to GAD and IA-2 (Sebastiani *et al.*, 2011). They found that T1D patients with autoantibodies have an increase in expression of miR-326 compared to antibody-negative T1D patients. miRNAs have been associated with β-cell death. Thus, the upregulation of miR-21, in the experimental model of non-obese diabetic (NOD) mice, decreased the levels of PDCD4, a tumour suppressor that induces cell death via the Bax family of apoptotic proteins. On the other hand, decreased levels of PDCD4 in βcells made them more resistant to death (Ruan *et al.*, 2011). Hezova *et al*. measured the expression of miRNAs in regulatory T cells in T1D patients (Hezova *et al.*, 2010). They found that miR-342 and miR-191 were downregulated, whereas miR-510 was upregulated. In NOD mice, a group found that levels of miR-29a/b/c were increased in the islets before diabetes manifestation and in islets exposed to proinflammatory cytokines (Roggli *et al.*, 2012). It was suggested that changes in the level of miR-29 could contribute to cytokine-mediated β-cell dysfunction in T1D. Numerous other miRNAs may be involved in the pathogenesis of T1D, as reviewed by Fernandez-Valverde (Fernandez-Valverde *et al.*, 2011).

1.9 The epigenetic and genetic architecture in gene regulation

There are different regulatory mechanisms that act on different levels to remodel the chromatin structure. For example, alterations in the chromatin structure can affect expression and repression of genes by enzymatic modification of core histones. The modifications of the histones result in conformational changes of the chromatin that alter the access of promoters to transcription factors. Alongside transcription factors, several cis-regulatory elements including enhancers, promoters, silencers and insulators, are crucial to the function of the genome.

An enhancer is a region of DNA that enhances transcription levels of a gene through the binding of transcription factors (Hardison and Taylor, 2012). There are more than a million enhancers, therefore, many more than there are genes. Because of this, a number of genes are regulated by the same enhancer, which may co-localize with CpGs (Ziller *et al.*, 2013). Also it has been suggested that lineage-determining transcription factors may define the epigenetic and transcriptomic states by selecting enhancer sites, facilitating binding of signal-dependent factors (Heinz *et al.*, 2013). Gene enhancers can be found upstream or downstream of genes and do not necessarily act on the closest promoter i.e. they can act as distant promoters. In order to do this, the DNA loops around bringing the specific promoter to the initiation complex; this enhancer-promoter loop has approximately 120 kilobases (de Laat and Duboule, 2013) (Figure 1.5). Enhancers may be accompanied by insulators, which are located between the enhancers and promoters of adjacent genes and can limit phenotypic gene expression despite genetic activation (Hardison and Taylor, 2012).

Genetic activation occurs following binding of transcription factors to DNA. A section of DNA is made available to transcription factors binding by unwinding of the chromatin with reduced nucleosome density, low DNA methylation and the availability of selected sites to cleavage by DNase enzymes (DNase hypersensitivity sites or DHSs) (Consortium, 2012). DHSs represent regions of transcriptionally active genome and there were approximately 2.9 million such DHSs identified by the Encyclopedia of DNA Elements (ENCODE) Consortium (Thurman *et al.*, 2012). With the interaction between regulatory elements and epigenetic marks, regulation of gene expression is more complex than once thought (Hansen *et al.*, 2006).

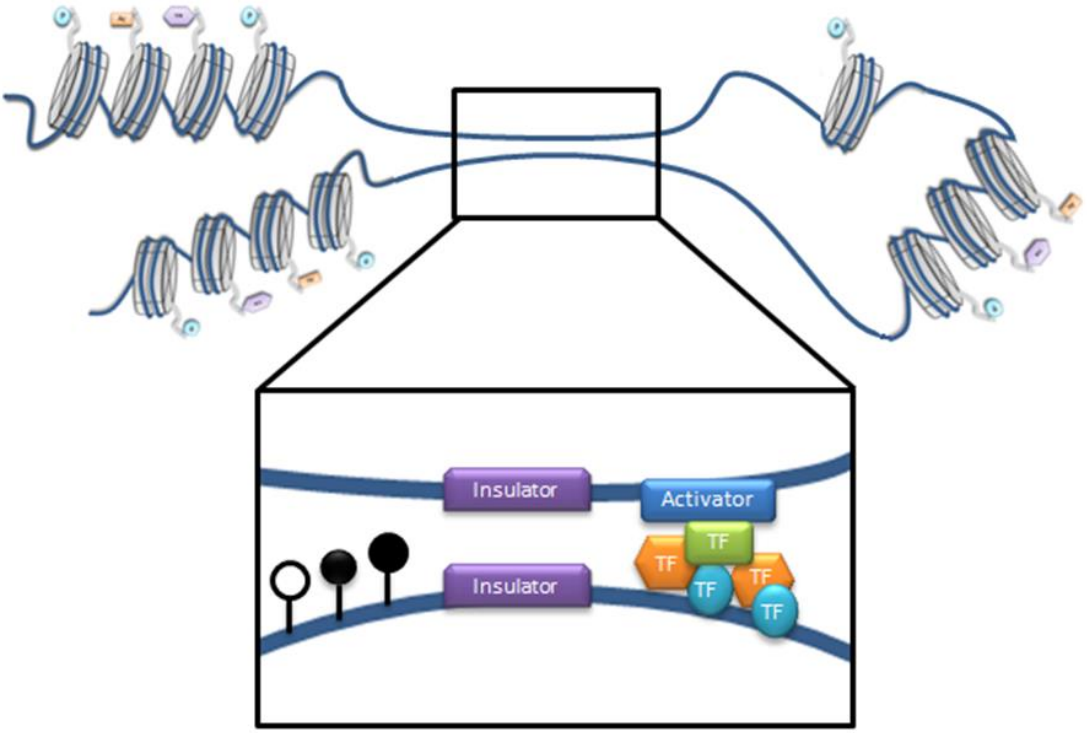


Figure 1.5. Interaction of regulatory elements. DNA is wrapped around a histone octamer and is joined to other nucleosomes by linker DNA. Post-translational modifications to the histone tails could occur by acetylation and methylation to alter the chromatin structure. An activator binds to an enhancer site, recruiting transcription factors (TF) to bind to the promoter site. This promoter site may be over 100kb away, therefore the DNA has to loop around. A pair of insulators interacts to block transcription occurring at a promoter site. A promoter site may have several CpG sites which could be methylated (black circles) or unmethylated (white circles).

1.10 Hypothesis

DNA methylation is associated with gene repression and is maintained through DNA replication. Significant methylation differences between MZ twin pairs discordant for T1D, will strengthen the hypothesis that epigenetics is involved in T1D aetiology. A preliminary study had found DNA methylation differences between MZ twin pairs discordant for T1D. The diabetic twins were all diagnosed under 20 years of age. This study will include twin pairs with adult-onset T1D, therefore hypothesising that the adult-onset T1D twins may have a larger difference in methylation compared to their healthy co-twin, due to more exposure to the environment.

In T1D patients, most β-cells are destroyed, releasing DNA into the circulation. It has been reported that in diabetic patients, unmethylated DNA from damaged β-cells is tissue-specific and is released into the circulation. As DNA methylation marks are stable, it was hypothesized that differences in methylation between MZ twin pairs discordant for T1D from circulating DNA can be detected and can be used as a biomarker for monitoring β-cell death.

1.11 Aims and objectives

The aim of the thesis was to study and investigate the role of DNA methylation in T1D. This involved generating whole-genome DNA methylation profiles of $CD14^{+}CD16^{-}$ monocytes, $CD4^{+}$ T cells, $CD19^{+}$ B cells and buccal cells from T1D-discordant MZ twin pairs to identify any differentially methylated regions. Another objective was to compare T1D-MVPs with those already identified in childhood-onset T1DM from the preliminary study. Also, an assay was developed to detect differences in DNA methylation in circulating DNA using serum samples from the same MZ twin pairs.

Chapter 2

Methods

2 Methods

The following methods were performed to identify possible methylation differences in MZ twin pairs discordant for T1D. Also, differences in unmethylated insulin DNA were investigated using patient serum. These studies were approved by the Northern and Yorkshire Research Ethics Committee (REC Reference Number: 06/MRE03/22). All individuals had given informed consent. Materials and equipment are listed in Appendix I.

2.1 Subject selection

T1D-discordant MZ twins were ascertained from The British Diabetic Twin Study. All diabetic twins selected, had T1D as defined by the American Diabetes Association (American Diabetes, 2009) and were referred through their physicians. All twin pairs were of European origin and monozygosity was confirmed by DNA fingerprinting. Neither twin was receiving drugs other than human insulin and statins in the index case. Diabetes was excluded from the non-diabetic twin by random whole blood glucose at testing less than 7.0 mmol/L (Hawa *et al.*, 1997). Blood glucose was measured using a Hemocue Glucose meter. The twins were also not conceived through artificial means. Average age at sampling was 43 years of age and the average age at diagnosis in the index twin was 19 years of age.

2.2 Testing for GAD65, IA-2 and ZnT8 autoantibodies

Radioimmunoassay was carried out to detect diabetes-associated autoantibodies to GAD65, IA-2 and ZnT8. Radioimmunoassay remains the "gold standard" assay to detect autoantibodies associated with T1D (Wenzlau and Hutton, 2013). The assays were carried out according to The Diabetes Antibody Standardization Program (DASP), which aims to standardise measurements of autoantibodies associated with T1D, between different laboratories (Schlosser *et al.*, 2010). Full length complementary DNA (cDNA) were transcribed and translated in the pCR11 (IA-2), pB 1882 (GAD65) and pCDNA3.1 (ZnT8) vectors and were then used to detect the autoantibodies in patient serum.

2.2.1 Lysate transcription and translation for GAD65, IA-2 and ZnT8

The transcription and translation method was performed using the TNT® Coupled Reticulocyte Lysate System from Promega. The TNT lysate reaction, using [^{35}S] methionine, consisted of the mix: 25 µL TNT rabbit reticulocyte lysate, 2 µL TNT reaction buffer, 1 µL amino acid mixture, 1 µL RNasin ribonuclease inhibitor, 1 µL TNT RNA polymerase SP6 (GAD65) or T7 (IA-2 and ZnT8), 2 µL DNA templates from GAD65, IA-2 or ZnT8, 4 µL [^{35}S] methionine (1,000 Ci/nmol at 10 Ci/mL) and 14 µL nuclease free water to make the final volume to 50 µL. The reaction mix was incubated at 30°C in the water bath for 1.5 hours. Following incubation, a NAP-5 column was washed five times with 1x Tris Buffered Saline tween (TTBS). The reaction mix was diluted with 200 µL 1x TTBS and was applied to the column. The column was washed with 400 µL 1x TTBS and a further 1 mL 1x TTBS was added to the column for elution. The first elute which was clear in colour, was discarded. The eluent, which contained haemoglobin, was collected. The excess radioactivity from the column was washed out with distilled water 25 times using an aspirator and then discarded. The translated protein was stored at -70°C until used for radioimmunoassay.

2.2.2 Radioimmunoassay for GAD, IA-2 and ZnT8 autoantibodies

Blood samples for the radioimmunoassay were collected in red top serum vacutainers and were left to clot at room temperature (RT) for at least 30 minutes. After the samples had clotted, the serum fractions were obtained by centrifugation, at 1,500 revolutions per minute (rpm) for 15 minutes at 20°C. These samples were split into three cryovials and frozen at -20°C. All samples were tested in duplicate, including positive and negative control standard sera.

The radioactivity was determined by precipitation with 10% trichloroacetic acid. For the immunoprecipitation in each assay, 50 µL of the antigen labelled with 35S methionine was incubated with 2 µL serum in 1x TTBS in a 96-well filter plate. The plate was incubated for 24 hours at 4°C on a plate shaker (180 rpm for five minutes, then 70 rpm overnight). Protein A Sepharose was pre-swollen in 1x TTBS overnight, then was spun at 1,300 rpm for three minutes at 20°C, followed by a quick wash with 1x TTBS. The supernatant was discarded and 15 mL 1x TTBS was added to the pellet and mixed well. With this, 1 mg (50 µL) Protein A Sepharose was added to

each well. The plate was incubated at 4°C in a shaker for one hour at 180 rpm and then washed with 1x TTBS (200 µL/well) using a vacuum eight times. The plate was blot dried using a paper towel and then left to air dry for approximately 20 minutes at RT. A plastic seal was placed over the plate and the radioactivity was determined by scintillation counting using a Micro-β-counter. The results were exported onto Microsoft Excel and values were compared to the World Health Organisation (WHO) units and laboratory standards. This determined whether the patient was positive or not for diabetes-associated autoantibodies.

2.3 DNA extraction from buccal cells

Buccal cell brushes provide another way of obtaining DNA and are another approach for collecting a sample from subjects in a non-invasive way. Buccal cells have been found to have six times as many hypomethylated regions compared with peripheral blood (Lowe *et al.*, 2013). Therefore, analysis of buccal cells may detect DNA methylation variation not picked up in other tissues such as peripheral blood. DNA was extracted from buccal cells using the Gentra Puregene Buccal Cell Kit from Qiagen according to manufacturer's instructions. Centrifugation was performed at 14,000 rpm at RT unless specified otherwise.

Per person, buccal cell samples were collected with four cytology brushes. The brush was placed into an eppendorf tube and the stick was cut off. Into the tubes, 600 µL Cell Lysis solution was added and stored at 4°C until DNA extraction of the samples. When ready to process, 2 µL Proteinase K was added to the tubes to degrade any proteins, inverted 25 times and incubated at 55°C overnight. The collecting brush was removed and 2 µL RNase A solution was added to prevent RNA contamination and mixed by inverting 25 times. The samples were incubated at 37°C for one hour and then placed on ice for one minute. 200 µL Protein Precipitation Solution was added to the tubes, vortexed for 20 seconds and incubated on ice for five minutes. The samples were spun for three minutes, placed on ice for five minutes and centrifugation was repeated. 600 µL isopropanol and 1 µL glycogen solution were dispensed into new eppendorf tubes. Supernatant from the samples were decanted into the new eppendorfs and mixed by inverting gently 50 times. The samples were spun for five minutes and supernatant discarded. The pellets were washed with 300 µL of 70% ethanol by spinning for one minute. Supernatant was discarded and left to air dry for at least 45 minutes. 30 µL DNA Hydration Solution was added to the tubes, vortexed for five seconds and then

incubated at 50°C for one hour. The samples from the same individual were combined at the end of the DNA extraction protocol. DNA concentration was determined by a Qubit instrument and visualised on a 2% agarose gel.

2.4 Immune cell isolation

Earlier studies have examined DNA methylation profiles in PBMCs in human diseases (Li *et al.*, 2010; Steegenga *et al.*, 2014). However, any disease-related signatures identified would just display differences in the relative abundance of individual cell types as each cell subset generates a unique methylation profile (Reinius *et al.*, 2012). Therefore it is important to isolate specific cell types to identify cell- or tissue-specific changes in DNA methylation to avoid confounding variables. This method involved isolating $CD19^+$ B cells, $CD14^+CD16^-$ monocytes and $CD4^+$ T cells for epigenetic profiling.

2.4.1 B cell isolation

Blood taken in heparin vacutainers was diluted 1:1 with Roswell Park Memorial Institute (RPMI) and left on the roller overnight. The diluted blood was carefully layered onto 12.5 mL Percoll 1.078g/mL in 50 mL Falcon tubes and centrifuged at 800 xg for 20 minutes with no brake, at 20°C. Percoll was used as it is a simple method for isolating PBMCs (de Almeida *et al.*, 2000). The sample had separated into different layers (Figure 2.1) where most of the plasma layer was removed before the PBMC ring was transferred to new 50 mL Falcon tubes. The tubes were topped up with phosphate buffered saline (PBS) and ethylenediaminetetraacetic acid (EDTA) and spun at 550 xg for eight minutes at 20°C. The cell pellets were pooled together and washed again. The cells were resuspended in 10 mL PBS and EDTA and counted using a haemocytometer. The tubes were topped up to 45 mL of PBS and EDTA and spun at 550 xg for six minutes at 4°C. $CD19^+$ B cells were then positively selected by magnetic-activated cell sorting (MACS). MACS separates different cell populations through the use of magnetic beads coated with antibodies against the surface antigen of interest. The cells were then passed through a column in a magnetic field and the cells of interest were retained in the column to obtain a pure cell population. MACS has been shown to obtain high purities of isolated cell populations, and subsequent use of magnetic beads on a single sample did not significantly affect gene expression (Lyons *et al.*, 2007). The cell pellet was resuspended in 80 µL PBS and EDTA per 10 million cells. Then 20 µL of CD19

microbeads was added per 10 million cells. This was left to incubate for 15 minutes at 4°C. The samples were washed with 2 mL of PBS and EDTA and spun at 300 xg for six minutes at 4°C. Whilst the samples were spinning, an LS column was washed with 3 mL PBS and EDTA through a pre-separation filter. The cells were resuspended in 500 µL of PBS and EDTA and applied to the column through the pre-separation filter. The flow-through was collected in another tube placed under the column. 3 mL of PBS and EDTA was added to the tubes to collect any remaining cells and added to the column three more times. The column was removed from the magnet and 5 mL of PBS and EDTA was added to the column. The purified cells were flushed out using the plunger. The flow-through and purified cells were spun at 300 xg for six minutes at 4°C. The purified cells were applied to another column to order to increase the purity. Whilst the cells were spinning, an MS column was washed with 500 µL of PBS and EDTA. The supernatant from the $CD19^-$ tube was decanted and left on ice to one side. The $CD19^+$ fraction was resuspended in 500 µL of PBS and EDTA and added to the MS column. 500 µL of PBS and EDTA was added to the tubes and then to the column three more times. The column was removed from the magnet and 2 mL of PBS and EDTA was added to the column to flush out the purified cells. The flow-through was discarded and purified cells counted.

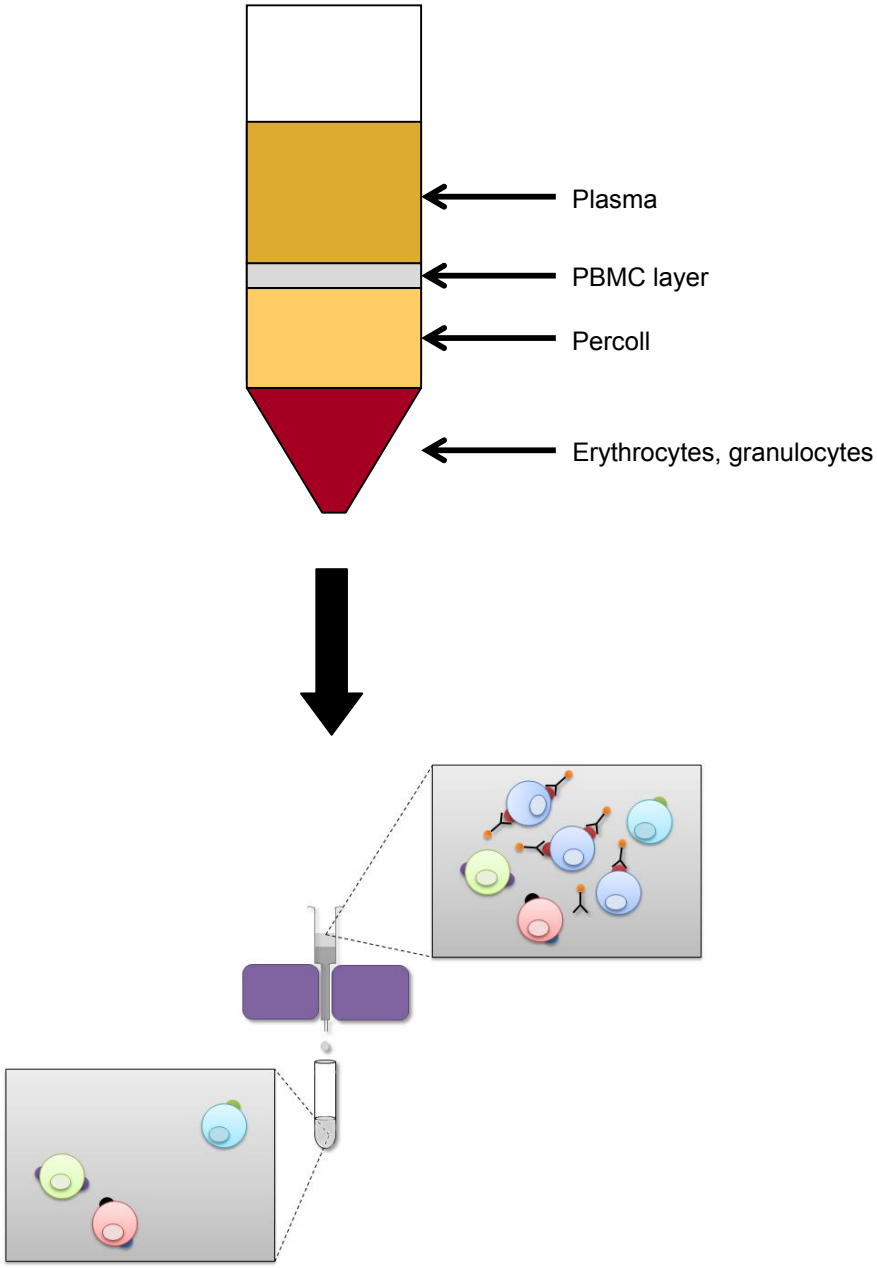


Figure 2.1. Separation and isolation of immune cells. The PBMC layer was first separated using Percoll. The cells of interest were labelled with magnetic microbeads and then loaded onto a MACS column which was placed in the magnetic field of a MACS Separator. The magnetically-labelled cells were retained within the column and the unlabelled cells flowed through. The purified cells were eluted with buffer when the column was removed from the magnetic separator. An aliquot of the purified cells was stained for flow cytometry analysis and the remainder of the cells were lysed for DNA or RNA extraction.

2.4.2 Monocyte isolation

The cell pellet left to one side from Section 2.4.1 (first $CD19^-$ fraction) was resuspended in 50 µL PBS and EDTA per 50 million cells. 50 µL of CD16 microbeads was added per 50 million cells. The tubes were incubated for 30 minutes at 4°C and washed by adding 2 mL of PBS and EDTA and then spun at 300 xg for six minutes at 4°C. Whilst the cells were spinning, an LD column was washed with 2 mL PBS and EDTA. The cell pellet was resuspended in 500 µL of PBS and EDTA and applied to the column. 1 mL of PBS and EDTA was added to the tubes to collect any remaining cells and added to column twice more. The flow-through was collected and a cell count was performed. After depletion of the $CD16^+$ cells, the sample was spun at 300 xg for six minutes at 4°C. The pellet was resuspended in 80 µL PBS and EDTA per 10 million cells. 20 µL of CD14 microbeads was added per 10 million cells and incubated for 15 minutes at 4°C. The cells were washed with 2 mL of PBS and EDTA and spun at 300 xg for six minutes at 4°C. Whilst the cells were spinning, an MS column was washed with 500 µL PBS and EDTA. The cells were resuspended in 500 µL of PBS and EDTA and applied to the column. 500 µL of PBS and EDTA was added to the tubes and then the column three times. The column was removed from the magnet and 2 mL of PBS and EDTA was added to flush out the purified cells. The flow-through was collected, counted and spun at 300 xg for six minutes at 4°C.

2.4.3 T cell isolation

For $CD4^+$ T cell isolation, the cell pellet from Section 2.4.2 was resuspended in 80 µL PBS and EDTA per 10 million cells. 20 µL of CD4 microbeads was added per 10 million cells. The tubes were incubated for 15 minutes at 4°C and washed with 2 mL of PBS and EDTA and spun at 300 xg for six minutes at 4°C. Whilst the cells were spinning, an MS column was washed with 500 µL PBS and EDTA. The cells were resuspended in 500 µL of PBS and EDTA and applied to the column. 500 µL of PBS and EDTA was added to the tubes and then to the column three times. The column was removed from the magnet and 2 mL of PBS and EDTA was added to the column to elute the purified cells.

Aliquots of purified cells ($CD14^+CD16^-$, $CD4^+$ and $CD19^+$) were taken for flow cytometry analysis and the rest were split equally into two 1.5 mL eppendorf tubes and spun down at full speed for two minutes. One of the pellets was resuspended in

200 µL PBS and 200 µL Buffer AL from the QIAamp DNA Blood Mini Kit, Qiagen, vortexed and stored at -20°C for later DNA extraction. The other pellet was resuspended in 300 µL RNAProtect, vortexed and stored at -80°C for later RNA extraction.

2.4.4 Optimisation of the immune cell isolation protocol

Initially, PBMCs extracted from the samples were split into three to try and positively select the three different cell types. However, this resulted in low cell numbers and therefore would not obtain enough DNA for downstream experiments. The harvested PBMCs were then split into two, one to sort for $CD19^+$ only and the other to sort for $CD14^+CD16^-$ and then $CD4^+$ cells. However, this also resulted in low cell numbers for $CD19^+$. As $CD19^+$ cells make up only 5-15% of lymphocytes (Fettke *et al.*, 2014), it was crucial to start with all the PBMCs.

Additionally, a second Percoll step was included in Section 2.4.1 to increase the purity of the cells and to reduce debris in the samples. However, this step was later removed as final cell numbers were low and subsequently, resulted in low DNA concentrations. The purity of $CD19^+$ cells dropped significantly, hence the addition of another column step for this cell type to increase the purity to over 95%.

2.5 Flow cytometry

Flow cytometry is a system for detecting cells by a laser beam. The analysis and differentiation of the cells is based on size and granularity by measuring the relative scattering and colour discriminated fluorescence of the particles. Flow cytometry can be used for different purposes; from sorting cells to determining purity of specific cells in a sample. All cells were analysed for purity.

2.5.1 Fluorescence compensation

Separate detectors in the flow cytometer measure the fluorescence from different fluorochromes. However, some fluorescence may be detected in the wavelength range that is assigned to a different flurochrome. Therefore, when two or more dyes are used to stain cells, the spectral overlap can be corrected by a process called compensation (Herzenberg *et al.*, 2006). Compensation estimates the percentage of

spectral overlap and subtracts it from the total detected signals resulting in an estimate of the actual amount of each dye.

Fluorescence compensation was performed with anti-mouse Ig compensation beads. Each FACS tube for each fluorochrome had anti-mouse Ig, κ particles, the negative control and the fluorochrome measured. The mixture can determine distinct positive and negative stained populations which can be used to set compensation levels.

2.5.2 Flow cytometry controls

In order to establish gating boundaries or to identify staining issues, isotype controls and fluorescence minus one (FMO) controls were used. Unstained cells were also analysed by flow cytometry (Appendix II, Figure 8.1).

2.5.2.1 Isotype controls

Purified cell populations were stained with mouse IgG isotype control antibodies. Isotype controls were used to assess the level of background staining. The antibodies were specific for keyhole limpet hemocyanin and as this protein is not expressed on human cells, it can be used as a negative control to distinguish specific from non-specific binding of mouse IgG1 fluorochrome–conjugated antibodies to human cells. The gating can then be used as a reference for future experiments to determine the specific cell population being investigated.

2.5.2.2 Fluorescence minus one controls

FMO controls are a way to determine gating cells of interest. In an FMO control set, all antibodies were used when staining cells except the antibody for which the threshold is to be determined (Herzenberg *et al.*, 2006). These controls will optimise the identification of the cells investigated and will also determine the cell population's autofluorescence (Appendix II, Figure 8.2).

2.5.3 Flow cytometry analyses

Single-cell suspensions were prepared from the purified cell populations and assessed for purity on the BD FACSCanto™ II. The data obtained from the BD FACSCanto™ II were analysed on FlowJo, a software package that can analyse flow cytometry data.

For surface staining, $CD14^+CD16^-$ cells were stained with 20 µL FITC conjugated monoclonal mouse anti-human CD14, clone MφP9, 20 µL PE conjugated monoclonal mouse anti-human CD16, clone B73.1/leu11c, 5 µL PerCP-Cy5.5 conjugated monoclonal mouse anti-human CD64, clone 10.1, 5 µL PE-CY7 conjugated monoclonal mouse anti-human CD45, clone HI30. $CD4^+$ cells were stained with 10 µL FITC conjugated monoclonal mouse anti-human CD4, clone M-T466. $CD19^+$ cells were stained 10 µL with PE conjugated monoclonal mouse anti-human CD19, clone LT19. Cells were incubated for 15 minutes at 4°C, washed with 2 mL PBS and EDTA and resuspended in 500 µL for FACS analysis. Expected overall percentage purity of the isolated cell populations was over 95%.

In a PBMC sample, there were many different populations of cells, in particular, $CD16^+$ cells (Figure 2.2A and B). For isolating the $CD14^+CD16^-$ monocytes, it was important to deplete the sample of $CD16^+$ cells. Flow cytometry analysis showed this to be effective (Figure 2.2C and D). The addition of two antibodies (CD64 and CD45) was used to also stain for monocytes (Figure 2.3A and B). The monocyte populations that were gated were then assessed for CD45 and CD64 markers shown in the middle panels. The purity of $CD14^+CD16^-$ cells can be seen in the right panels. As expected, there was a higher purity of $CD14^+CD16^-$ cells after the CD16 depletion and CD14 purification steps. As shown, the purity of monocytes prior to $CD16^+$ cells depletion was 93.2%, compared to after the depletion of $CD16^+$ cells at 97.4%. $CD4^+$ T cells had the highest purity at 99.1% (Figure 2.3C) and $CD19^+$ cells with 95.5% (Figure 2.3D).

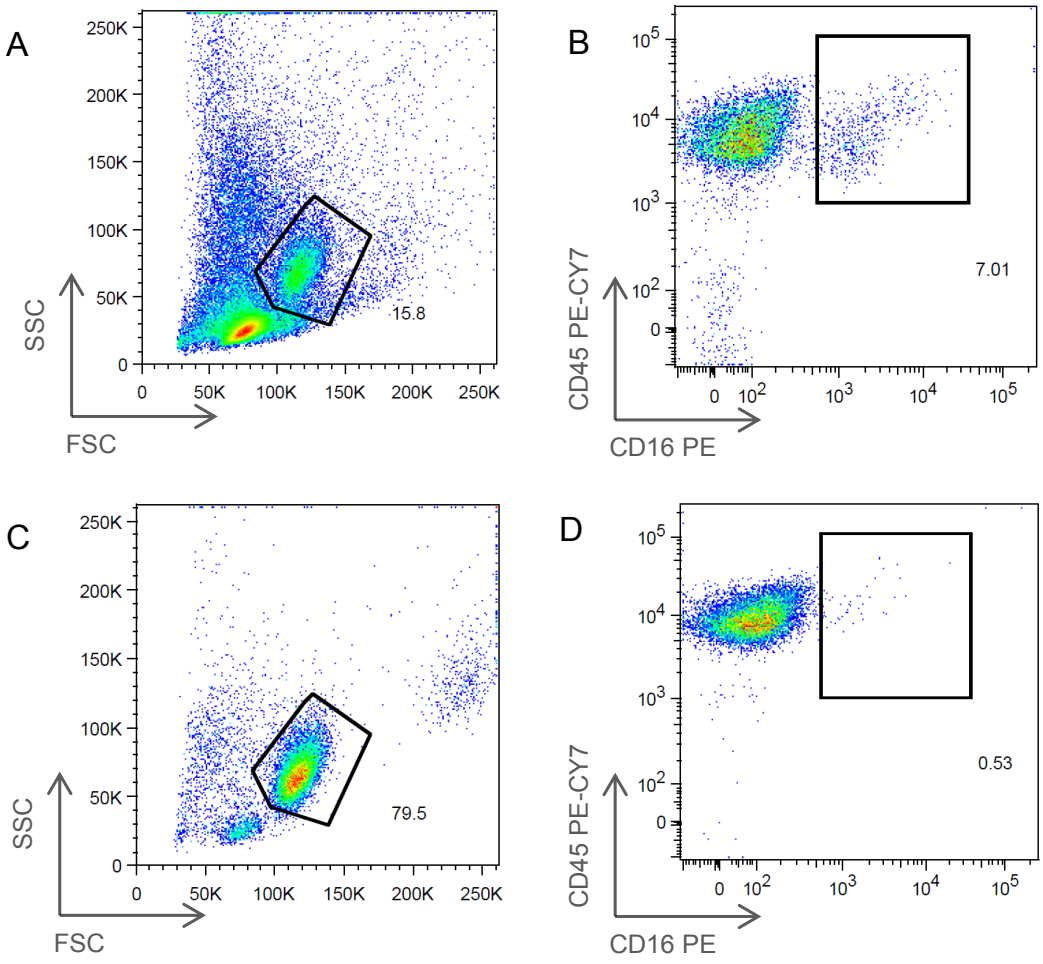


Figure 2.2. Depletion of CD16$^+$ cells from PBMC samples. Flow cytometry analysis of PBMC staining before magnetic beads separation (A). Gating for monocytes, this resulted in 15.8% of the PBMC sample. Within this gate, 7.01% cells expressed CD16 (B). A total of 10,000 events were recorded in the gates. Monocyte staining after CD16 depletion and positive selection for CD14 (C) resulted in a purer population of monocytes in the PBMC samples at 79.5%. Within this gate, 0.53% of cell expressed CD16, showing that the depletion of CD16 using MACS was effective (D).

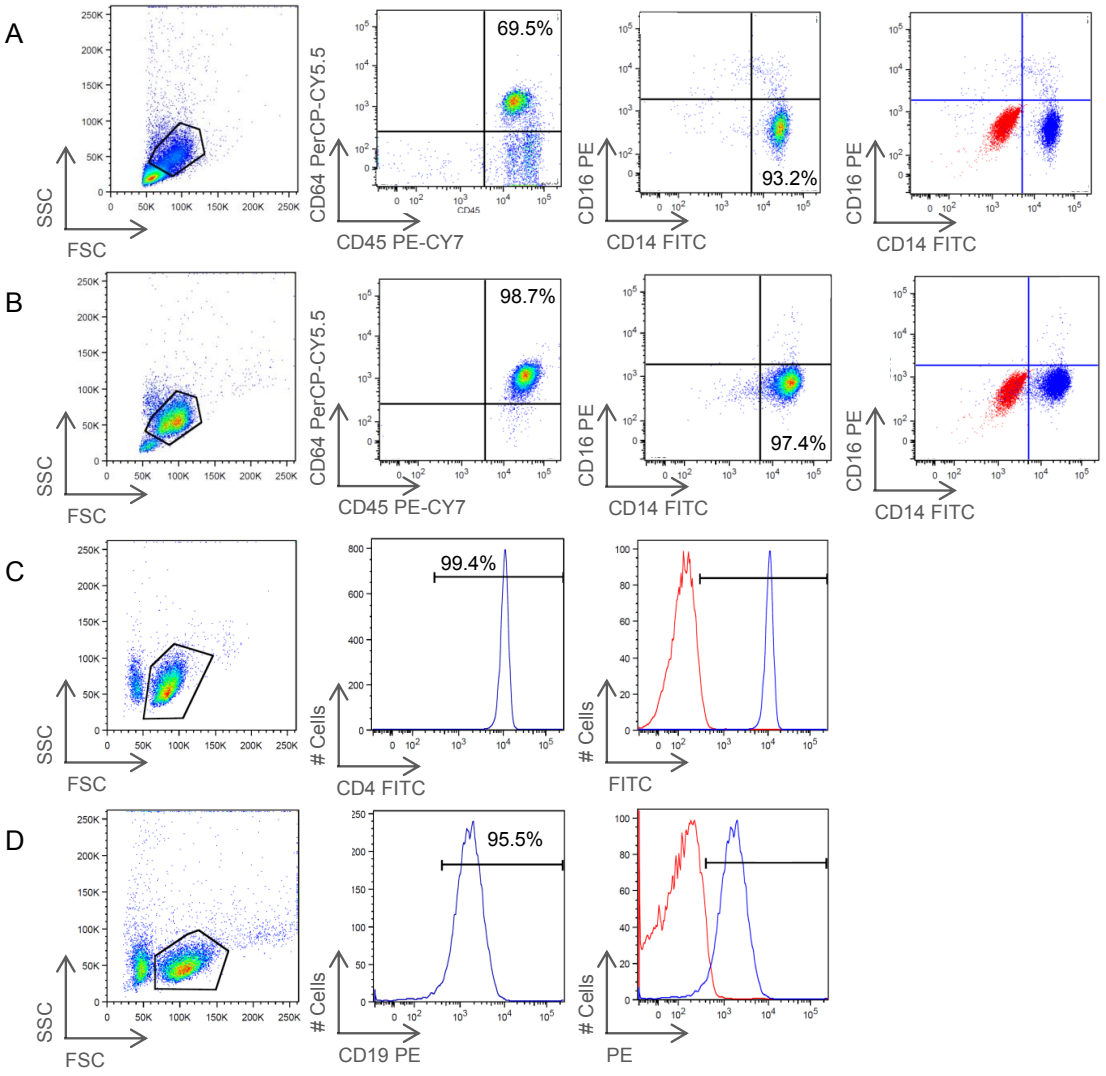


Figure 2.3. FACS analyses of selected cells. FlowJo analyses of PBMCs before the CD16 depletion step, gated for monocytes (A). FlowJo analyses of $CD14^{+}CD16^{-}$ sample after the CD16 depletion and $CD14^{+}$ purification steps (B). Flowjo analysis of $CD4^{+}$ cells (C). $CD4^{+}$ cells were stained with antibodies labelled with FITC. D) Flowjo analyses of $CD19^{+}$ cells. 10,000 events were recorded within the gates in all the FSC/SSC plots. Mouse anti-IgG antibodies was used to stain the cells to establish a negative control (red).

2.6 Staining of isolated cells

Further staining of the purified cells enabled further determination of whether the right cell type was isolated. This was performed using Romanowsky staining. Romanowsky stains combine methylene blue and eosin dyes to stain the nucleus and cytoplasm (Jorundsson *et al.*, 1999). The dyes bind to acidic nuclei and result in a blue to purple colour.

The slides were pre-wetted with 50 µl of PBS and EDTA to wet the pad paper as the sample volume was too small. 25,000 cells in 100 µl of PBS and EDTA was added to the funnels in the cytospin clips and placed into the cytospin centrifuge. After spinning, the slides were left to air dry for approximately 20 minutes. The slides were immersed into the Rapid Romanowsky Solution A for 30 seconds. The slides were removed and excess tapped off, immersing the slides into Solution B (Eosin Y) for 30 seconds. The slides were removed and immersed into Solution C (Methylene Blue) for 30 seconds. The slides were removed and washed under running tap water to remove excess stain. The cytospins were air dried for 1 hour before mounting with DPX mounting medium. After using a Romanowsky stain kit, the slides were viewed with a light-emitting diode (LED) microscope (Figure 2.4).

2.7 DNA extraction from cells

Genomic DNA was extracted from the lysed purified cell pellets using the QIAamp DNA Blood Mini Kit according to manufacturer's instructions. The Qiagen protease and Buffers AW1, AW2 were prepared following manufacturer's instructions. All centrifugation steps were performed at 8,000 rpm at RT unless stated otherwise.

Frozen samples for DNA extraction from Section 2.4.3 were thawed at RT then 20 µL Qiagen Protease was added to break down any proteins. The samples were mixed by pulse-vortexing for 15 seconds and then incubated at 56°C for 10 minutes. 200 µL of 100% ethanol was added to the samples, mixed well and applied to the QIAamp Mini spin column. The columns were spun for one minute with the collection tube then discarded, placing the column into a new 2 mL collection tube. To wash the columns, 500 µL Buffer AW1 was added and the tubes were spun for one minute. The columns were placed into a new 2 mL collection tube and were washed again with 500 µL Buffer AW2. The tubes were spun at 14,000 rpm for three minutes. The columns were placed in new 1.5 mL tubes and centrifuged at 14,000 rpm for one minute to remove any remaining buffer left in the columns. The columns were transferred to a new 1.5 mL tube and to elute the DNA, 150 µL Buffer AE was added to the column membranes. The tubes were centrifuged for one minute and the DNA was recovered and measured using the Qubit instrument according to manufacturer's instructions. The samples were also run on a 2% agarose gel.

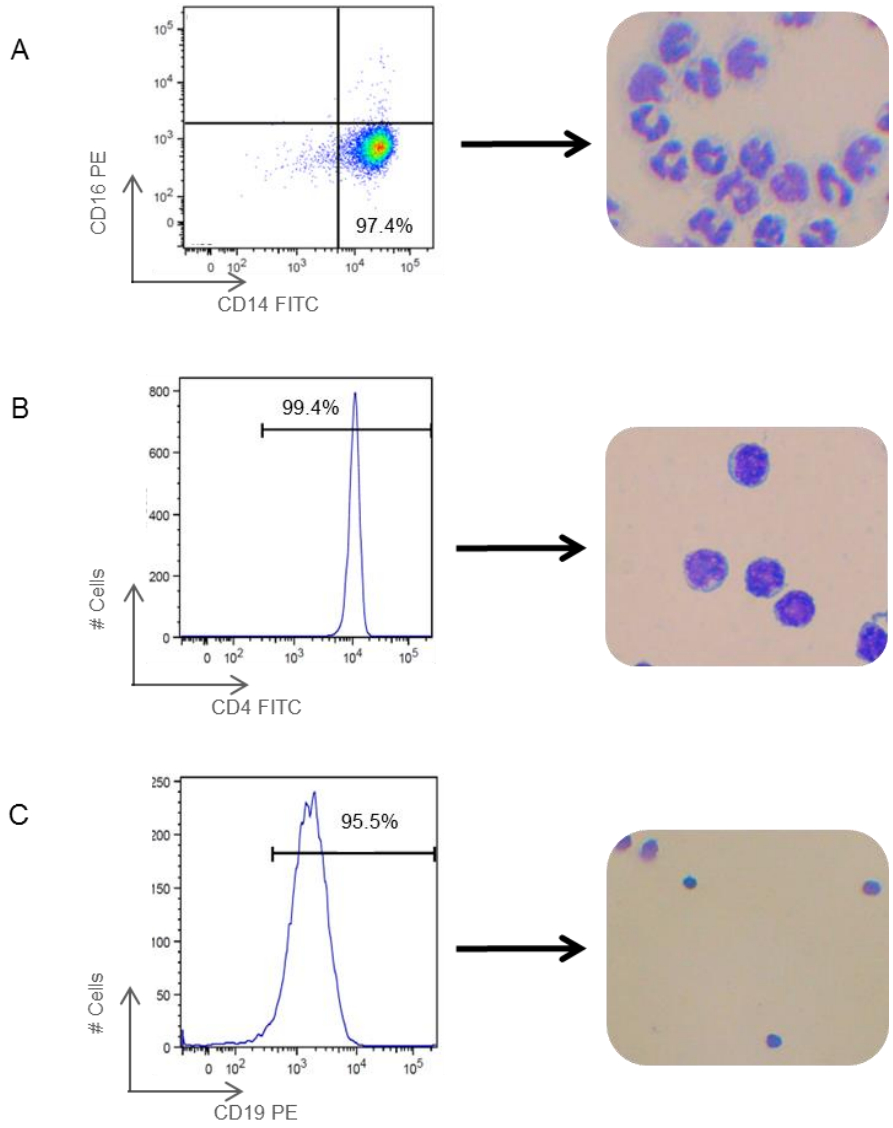


Figure 2.4. Staining of isolated cells using Romanowsky stain. Once the purified cells had been isolated by MACS, the cells were spun onto a slide and stained with a Romanowsky stain kit. This was to confirm that the cells of interest were correctly purified in comparison to flow cytometry analyses. Monocytes present with a horse-shoe shaped nucleus (A). T cells have a single large nucleus, which takes up most of the cell leaving a small amount of cytoplasm (B). B cells also have large nuclei, similar to T cells (C).

2.8 Agarose gel electrophoresis

DNA samples were run on a 2% agarose gel to assess the integrity of the samples. The 2% gel was prepared by dissolving 1.4 g agarose completely in 70 mL Tris/Borate/EDTA (TBE) buffer by mixing and heating. The gel was left to cool to approximately 55°C and 5 µL of ethidium bromide was added to the gel. The gel had to be cooled down before the addition of ethidium bromide to prevent toxic vapours as it is a mutagen. Ethidium bromide is a fluorescent dye that intercalates between bases of nucleic acids. This allows the visualisation of the samples in the gel. The tank was set up and the gel was slowly poured into the tank with a comb. The gel was left at RT to solidify. Whilst waiting for the gel to solidify, the samples were prepared. The samples and loading buffer were mixed in the ratio of 3:1. Once the gel had solidified, the comb was removed and the tank was filled with 1x TBE buffer. The samples were loaded into the wells and the gel was run at 70V for approximately 20 minutes. Once the run was completed, the gel was viewed in the UV Transilluminator.

2.9 DNA extraction from cord blood

In addition to the current samples, cord blood was analysed. Normally, blood samples from new-borns are taken in the form of dried blood spots, to screen for various diseases. Guthrie cards are useful with addressing issues such as causality of a disease, particularly in EWAS. Profiling the methylome using Guthrie cards may identify epigenetic variants that may indicate that it is disease casual as they appear before disease. DNA methylation variation has been shown to be stable for at least three years, indicating that there may be disease-related epigenetic marks that could be detected at birth (Beyan *et al.*, 2012a). The dried blood spots (DBSs) consisted of cord blood dotted onto card, from 99 new-borns from a cohort in Sweden. 50 of those new-borns went on to develop T1D and the other 49 did not. The DBSs were processed using the GenSolve DNA recovery kit from Labtech, as per the manufacturer's instructions. The steps that required centrifugation were performed at 8,000 rpm at RT unless specified otherwise.

The DBSs were stored in eppendorf tubes at 4°C. First, 620 µL Recovery Solution A/Protease mix was added to the tubes. The samples were placed in a thermomixer heated to 65°C and were set to shake at 14,000 rpm for one hour. The tubes were spun down briefly to collect any liquid off the cap. 20 µL Recovery Solution B was

added to new collection tubes and spin baskets were inserted. The DBSs from the previous step were transferred into the spin baskets. The tubes were then spun at 14,000 rpm for two minutes at RT. The spin baskets were then discarded and the remaining samples were taken through to DNA extraction using the QIAamp DNA Blood Mini Kit from Qiagen. 600 µL of 100% ethanol was added to each tube and vortexed. The samples were loaded onto spin columns and spun for one minute. The columns were placed into new 2 mL collection tubes and 500 µL AW1 Buffer was added. The tubes were spun for one minute and the spin columns were transferred into new 2 mL collection tubes. 500 µL AW2 Buffer was added and the tubes were spun at 14,000 rpm for four minutes at RT. The columns were placed into new 1.5 mL collection tubes and spun at 14,000 rpm for 1.5 minutes at RT. DNA was eluted by adding 100 µL AE Buffer to the spin columns. The samples were incubated at RT for five minutes and then spun for one minute. Eluted DNA was measured using the Qubit instrument, visualised on a 2% agarose gel and stored at -20°C.

2.10 Illumina Infinium HumanMethylation450K BeadChip

DNA methylation plays an important role in regulating gene expression. One approach to detect differential methylation patterns in the genome has been developed by Illumina. The Illumina Infinium HumanMethylation450K BeadChip (Illumina450K) involves bisulfite converted DNA hybridized to the beadchip, creating a methylation profile detailing over 485,000 CpG sites per sample at single-nucleotide resolution (Sandoval *et al.*, 2011). The beadchip can analyse up to 12 samples per array and only requires 500 ng of bisulfite-converted DNA per sample. Compared to its predecessor, the Illumina Infinium HumanMethylation27K (Illumina27K), the current beadchip has over 17-fold CpG sites and covers 99% of RefSeq genes, 96% of CpG islands and coverage in island shores (Touleimat and Tost, 2012). Additional regions covered include CpG sites outside of CpG islands, non-CpG methylated sites identified in human stem cells, CpG islands outside of coding regions and miRNA promoter regions. The Illumina450K BeadChip also retains approximately 90% of content contained on the Illumina27K BeadChip. The array has two bead types for each CpG site, Infinium I and Infinium II (Figure 2.5).

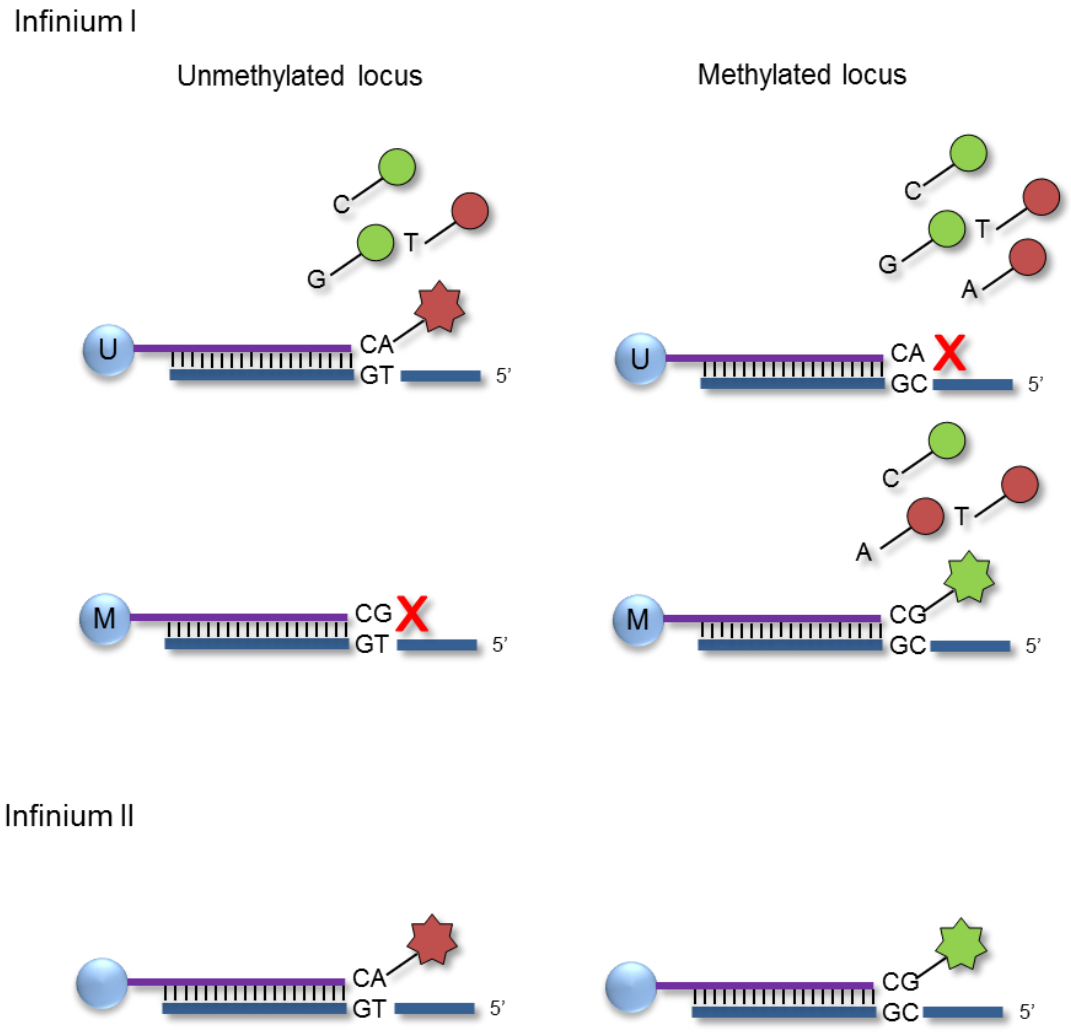


Figure 2.5. Illumina Infinium HumanMethylation450K BeadChip profiling technology. The Illumina450K BeadChip consists of two assays. The Infinium I assay interrogates each CpG using two bead types, methylated (M) and unmethylated (U) beads. If hybridization of an unmethylated locus occurs on the U bead, single-base extension with labelled nucleotides is performed. If a methylated locus hybridizes to the M bead, no single-base extension will occur, and vice versa. The Infinium II assay interrogates each CpG using a single bead type. Single-base extension always occurs at the end of the probe, therefore, methylation state is detected by the addition of a labelled 'G' or 'A' nucleotide. Following this, the BeadChip will then be fluorescently stained and scanned to measure the intensity of the beads corresponding to the unmethylated and methylated CpG sites.

2.10.1 Infinium Illumina450K BeadChip sample preparation

The Infinium Illumina450K assay combines bisulfite conversion of genomic DNA and array technologies to determine the methylation state of over 485,000 CpG sites (Figure 2.6).

2.10.1.1 Preparation of bisulfite-converted DNA

After DNA extraction of the samples, 500 ng of genomic DNA of each sample were prepared into a total volume of 45 µL and then loaded into 96-well plates. The samples were randomised to avoid batch effects. Batch effects can display different behaviour across conditions which are unrelated to biological variable in a study (Leek *et al.*, 2010). The automated BeadChip array was performed at the Genome Centre at University College London, UK. The samples were then treated with sodium bisulfite using EZ-96 DNA Methylation MagPrep kit according to the manufacturer's instructions. Sodium bisulfite treatment of DNA enables researchers to investigate DNA methylation. During sodium bisulfite treatment of the samples, unmethylated cytosines are converted to uracils; whereas methylated cytosines remain unchanged, therefore complete bisulfite conversion is crucial to identifying correctly the unmethylated or methylated fraction of the genome. After treating the DNA with sodium bisulfite, the samples were amplified. During amplification, unmethylated cytosines produce a T whereas the protected methylated cytosines retain the C (Figure 2.7). These products are then hybridized to custom microarrays that contained probes to discriminate converted versus unconverted cytosines at the CpG site of interest, thereby providing a readout of the original methylation state at that CpG site (Harris *et al*., 2010) (Figure 2.8).

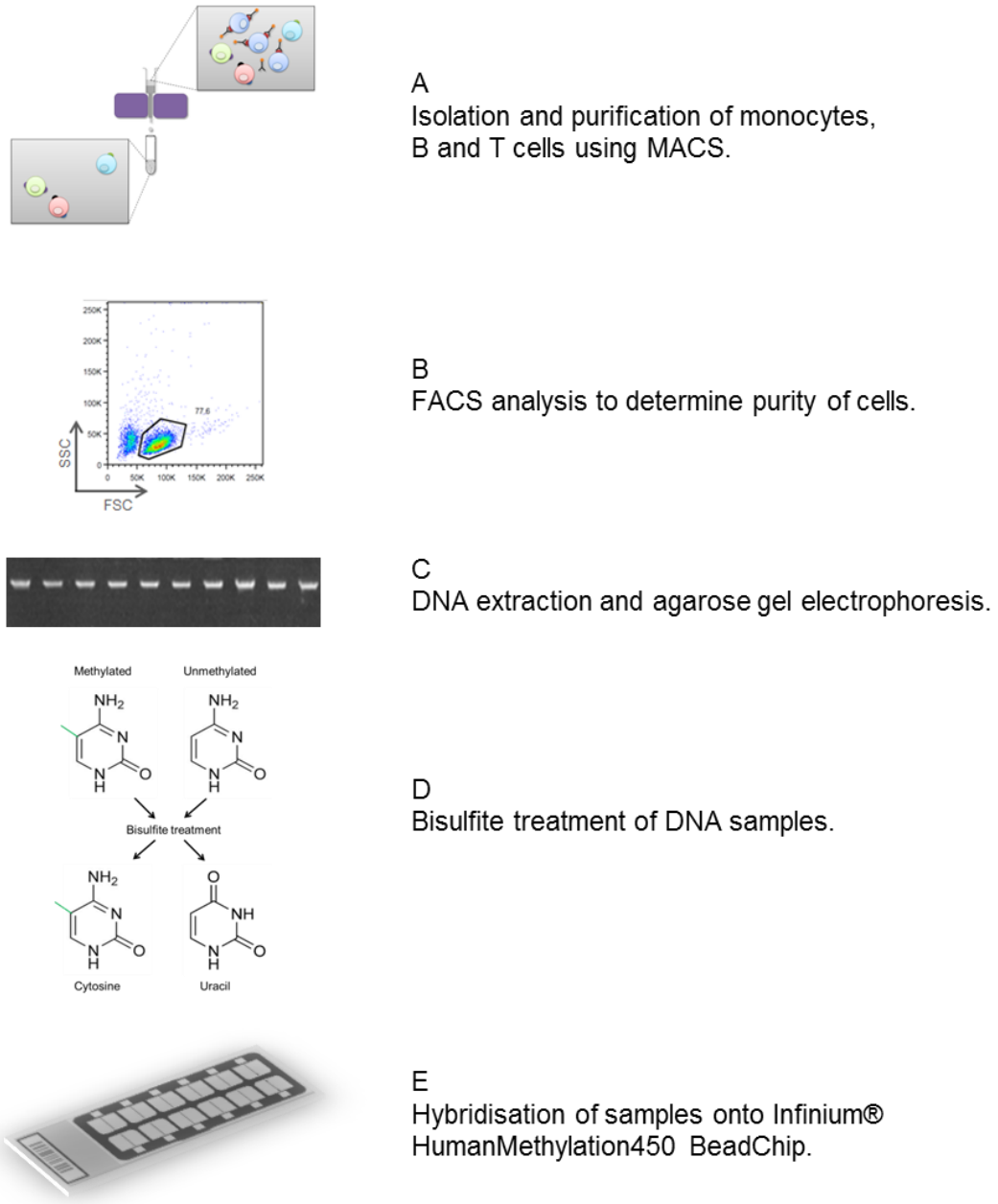


Figure 2.6. Summary of sample preparation for the Illumina450K BeadChip array. Cells of interest were separated and isolated using MACS (A). Flow cytometry was used to determine the purity of the cells isolated (B). DNA was extracted from each cell type and visualised on a 2% agarose gel to assess DNA integrity (C). DNA was treated with sodium bisulfite. Methylated cytosines remain protected whereas unmethylated cytosines in CpG sites convert into uracils (D). The bisulfite converted DNA was then hybridized onto an Illumina Infinium HumanMethylation450K BeadChip and analysed (E).

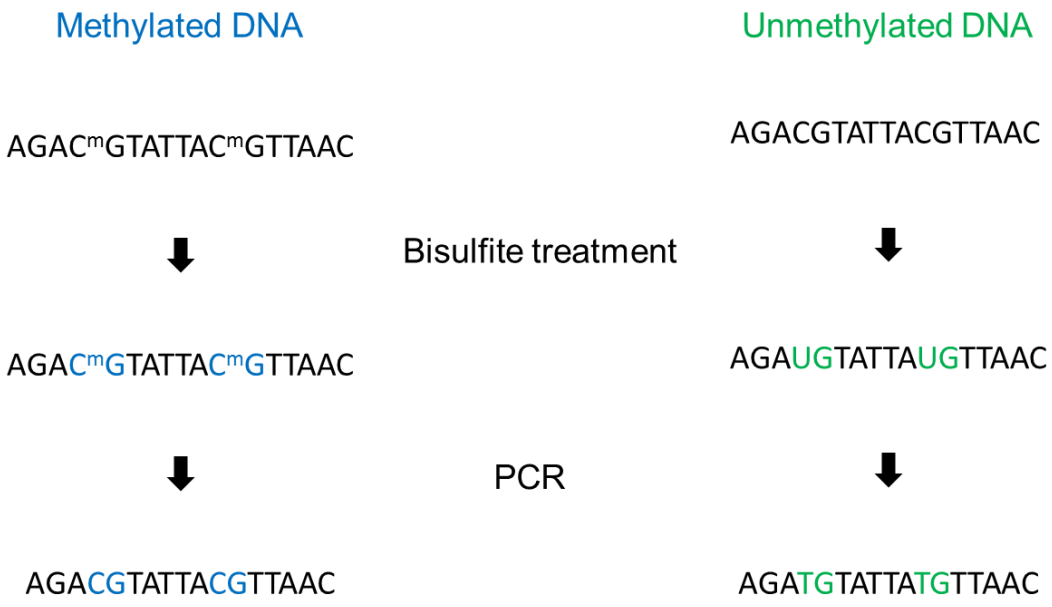


Figure 2.7. Bisulfite conversion of cytosines. During bisuflite conversion, the unmethylated cytosine is deaminated to produce uracil. Methylated cytosines are protected from the conversion to uracil. The DNA will then be amplified, during the amplification step, uracils are amplified at thymines whereas methylated cytosines get amplified as cytosines. The DNA sequence was then analysed, determining the positions of unmethylated and methylated cytosines.

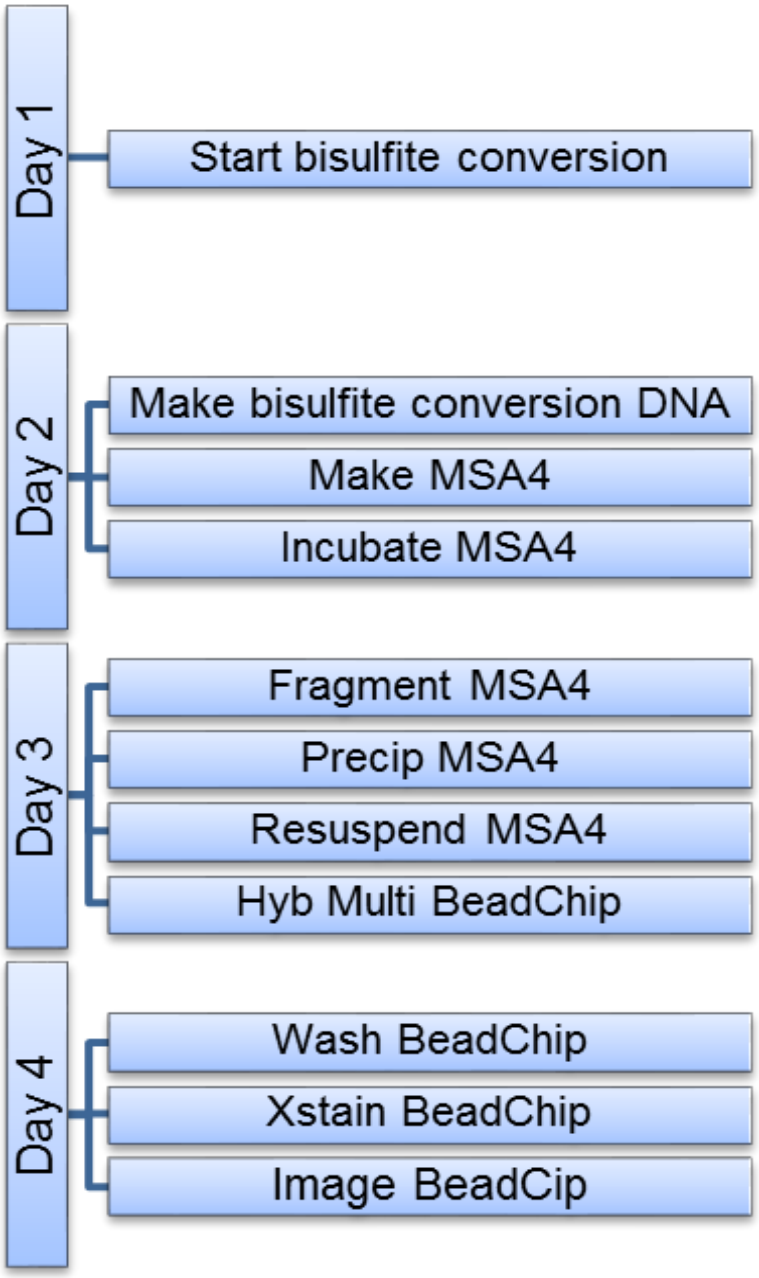


Figure 2.8. Summary of the automated Illumina450K protocol. Day 1 involves quantifying DNA, then the bisulfite conversion process begins, leaving it to incubate for 17-18 hours. On Day 2, the bisulfite conversion process was completed, creating MSA4 plate. MSA4 was left to incubate for 20-24 hours. On Day 3, MSA4 was fragmented and precipitated for 2 hours. MSA4 was resuspended for one hour, hybridized to the multi BeadChip and was left to incubate for 16-24 hours. On Day 4, the BeadChip was washed and stained, ready for it to be scanned. The image and data files were then transferred for bioinformatic analyses.

2.11 Bisulfite sequencing of T cells

Bisulfite sequencing (BS-seq) involves treatment of DNA with sodium bisulfite followed by DNA sequencing. BS-seq currently remains the "gold standard" technique for detection of 5-methylcytosine (5mC), providing the highest level of coverage and resolution, negligible bias towards CpG dense regions and a direct readout of non-CpG methylation (Rakyan *et al.*, 2011b). Data from BS-seq can be used to compare against Illumina450k data, as BS-Seq provides a more comprehensive coverage of the CpG sites in the genome.

The library preparation process was performed in the following stages: Sonication and clean-up of DNA, DNA end repair, adaptor ligation and purification, final repair, bisulfite conversion and amplification and purification. The library preparation was performed using the Ovation Ultralow Methyl-Seq Library Systems from Nugen.

2.11.1 Sonication and clean-up of DNA

Working aliquots of 300 ng genomic DNA was sonicated using a diagenode bioruptor resulting in DNA fragments of approximately 250 bp. The size was checked by running the samples on the bioanalzyer and then cleaned up using Ampure beads according to manufacturer's instructions. 1.8 volume of Ampure beads was mixed with 10 µL DNA and mixed thoroughly by pipetting. The samples were incubated at RT for 10 minutes. The tubes were placed on the magnet and left to stand for five minutes. The cleared solution was aspirated from the tubes and discarded. 200 µL of 70% ethanol was added to each tube and incubated for 30 seconds at RT. The ethanol was aspirated and another wash was repeated. The beads were left to air dry for 10 minutes to ensure that the ethanol had evaporated. 14 µL of 1x Tris-EDTA (TE) buffer was added to the tubes, mixed and was left to incubate at RT for one minute. The eluent was transferred to a new 0.2 mL tube.

2.11.2 DNA end repair

After sonication, the DNA fragments comprised of 3' or 5' overhangs. This step converted the overhangs into blunt ends. The master mix for one reaction contained 2 µL ER1 Buffer Mix, 0.5 µL ER2 Enzyme Mix and 0.5 µL End repair Enhancer. 3 µL of the master mix was added to each sample. The samples were incubated in a

thermal cycler for 30 minutes at 25°C, followed by 10 minutes at 70°C and held at 4°C.

2.11.3 Ligate adapters to DNA fragments

This step ligated adapters to the ends of the DNA fragments, preparing for hybridization to a flow cell. 3 µL of the L2 Ligation Adapter Mix was added to each sample. A master mix for a single reaction consisted of 4.5 µL D1 water, 6 µL L1 Buffer Mix and 1.5 µL L3 Enzyme Mix. 12 µL of the master mix was added to the samples and mixed thoroughly. The samples were incubated in a thermal cycler for 30 minutes at 25°C, followed by 10 minutes at 70°C and hold at 4°C. Post-ligation purification followed with the addition of 45 µL of the bead suspension and was processed as described in Section 2.11.1. The final elution volume was 15 µL. The final repair step involved preparing a master mix, for each reaction combining 4.5 µL Buffer FR1 and 0.5 µL Buffer FR2. 5 µL of the Final Repair Master Mix was added to each sample tube. The tubes were placed in the thermal cycler and were programmed to run at 60°C for 10 minutes, holding at 4°C.

2.11.4 Sodium bisulfite treatment of DNA samples

Bisulfite treatment of DNA was carried out using the EpiTect Bisulfite Kit from Qiagen according to manufacturer's instructions. Centrifugation was at maximum speed for one minute unless specified otherwise.

The 20 µL product of the Final Repair reaction was inputted directly into the bisulfite conversion kit. In the total reaction volume of 140 µL, 15 µL of DNA Protect Buffer and 85 µL Bisulfite Mix was added to the DNA, topping up with 20 µL RNase-free water. The bisulfite conversion was performed in a thermal cycler and was programmed to run at 95°C for five minutes, 60°C for 20 minutes, 95°C for five minutes, 60°C for 20 minutes and then hold at 20°C. The reactions were transferred to new 1.5 mL tubes and 560 µL of Buffer BL was added to each sample. The mixture was transferred to a spin column and then was spun. The flow-through was discarded and the column was placed back into the collection tube. 500 µL Buffer BW was added to each column and spun. The flow-through was discarded and column was placed back into the collection tube. 500 µL Buffer BD was added to each spin column and incubated at RT for 15 minutes. The column was spun. The flow-through was discarded and column placed back into the collection tube. 500 µL

BL was added to the column and spun. The column was placed into a new 2 mL collection tube and spun again to remove any residual liquid. The columns were then placed into new 1.5 mL tubes and then incubated at 56°C for five minutes. The column was placed into a new 1.5 mL tube and 20 µL Buffer EB was added to the columns. The purified DNA was eluted by spinning for one minute at 15,000 xg.

2.11.5 Polymerase chain reaction and clean-up of DNA

Polymerase chain reaction (PCR) is a technique which amplifies small amounts of DNA by a million-fold (Garibyan and Avashia, 2013). In a typical PCR reaction, DNA polymerase synthesizes a new DNA strand from the template strand. It does so with the help of primers, which are short single-stranded pieces of DNA that anneals to its complimentary sequence on the template once denatured. The DNA polymerase adds a nucleotide from the primers to create the new strand with deoxynucletide triphosphate (dNTPs). This in turn extends the annealed primer and amplifies the region of interest.

Following sodium bisulfite treatment of the DNA samples, PCR amplification was performed with primers that anneal to the ends of the adapters (Appendix I, Table 8.1). The master mix was prepared by combining 5 µL P2 and 25 µL P3 for each reaction. Then on ice, 30 µL of the master mix was added to each sample. The tubes were placed on the thermal cycler and programmed to run at 95°C for two minutes, eight cycles of 95°C for 15 seconds, 60°C for one minute and 72°C for 30 seconds, then hold at 10°C. The PCR products were then purified using the Agencourt Beads as described in Section 2.11.1 starting with 50 µL of beads and eluting with 20 µL of 10 mM Tris buffer, pH 8. The samples were run on the Bioanalyzer High Sensitivity DNA Chip 1000 to see whether the inserts were of the correct size. The libraries were then sent off to another research institute for sequencing.

2.12 Analyses of methylation data

After the samples had been sequenced, the data needed to be analysed and interpreted. There are different software tools for performing this (Bock, 2012). R is a software programming language that provides statistical and graphical techniques. Within R, graphical presentation of methylation data, such as clustering or principal component analysis, can be performed. There are different pipelines that are implemented in R for analysing methylation data.

For this study, the Chip Analysis Methylation Pipeline (ChAMP) pipeline (Morris *et al.*, 2014) was applied. Raw Intensity Data files (IDAT) files were imputed into R and with ChAMP, quality control and normalization was performed. The program also carries out batch effect analysis, SNP flagging and segmentation of MVPs into biologically relevant DMRs. Normalization is especially important, as the Illumina450k platform combines two different assays, Infinium I and Infinium II (Bibikova et al., 2011; Sandoval et al., 2011). BS-seq data were analysed using the *bsseq* package on the R platform (Hansen *et al.*, 2012).

2.13 Detection of β-cell death using real time PCR

In T1D, 80-90% of β-cells are destroyed at diagnosis (Notkins and Lernmark, 2001; Cnop *et al.*, 2005). DNA methylation differences have already been studied using patient serum samples in cancer (Anker *et al.*, 2001) and recently in T1D (Akirav *et al.*, 2011; Husseiny *et al.*, 2014). This section describes the development of a real-time PCR assay to detect unmethylated CpG sites in the insulin promoter region.

2.13.1 Mini-preparation of plasmid DNA

A pGL4 plasmid with the insulin promoter DNA sequence already inserted, was purchased from Addegene and deposited by Kevin Ferreri's laboratory (Figure 2.9A) (Kuroda *et al.*, 2009). The promoter region was targeted as it was shown to have tissue-specific unmethylated CpG sites in the β-cells compared to other tissues such as kidney, liver and colon (Husseiny *et al.*, 2014). The plasmid was then used as a positive control and to normalise the assay and to prepare a dilution series to test the amplification efficiency.

2.13.1.1 Preparation of bacterial plasmid DNA

The plasmid had already been transformed into DH5alpha (*E.coli*) bacteria and was received in a stab culture. In order to grow the transformed plasmid, an agar plate was made with Lysogeny broth (LB) agar (Invitrogen, UK). The agar medium was poured into sterile Petri dishes and was left to set. With disposable inoculating loops, bacteria from the stab culture were streaked across the plate to obtain single colonies. The plates were left overnight at 37°C. Single colonies were picked with a disposable tip and placed into a 50 mL Falcon tube with 5 mL LB medium and were left overnight shaking at 1,500 rpm at 37°C. 500 µL of the bacterial growth was added to 500 µL 50% glycerol in a 2 mL cryovial and stored at -80°C for long-term storage.

2.13.1.2 Plasmid DNA extraction

The rest of the bacterial growth was spun at 2000 xg for five minutes. The supernatant was decanted and pelleted bacterial cells were processed with the GeneJET Plasmid Miniprep kit according to manufacturer's instructions. The pellet was resuspended in 250 µL of Resuspension Solution and was transferred to a microcentrifuge tube. 250 µL of the Lysis Solution was added and mixed thoroughly

by inverting the tube until the solution became viscous and slightly clear. 350 µL of the Neutralization Solution was added and mixed immediately and thoroughly. The tubes were centrifuged for five minutes to pellet cell debris and chromosomal DNA. The supernatant was transferred to a GeneJET spin column and spun at 14,000 rpm for one minute. The flow-through was discarded and the column was placed back into the same collection tube. 500 µL of the Wash Solution was added to the column and was spun for one minute at 14,000 rpm at RT. The flow-through was discarded and the column was placed back into the same collection tube. The wash procedure was repeated with an additional spin for remove residual wash solution. The spin column was placed into a new 1.5 mL microcentrifuge tube and 50 µL of the Elution Buffer was added to the centre of spin column membrane to elute the plasmid DNA. The columns were incubated for two minutes at RT and centrifuged for two minutes. The column was discarded and the purified plasmid was stored at -20°C.

2.13.1.3 Sequencing of plasmids

To be certain the plasmids contained the right DNA insert, the plasmids were also sequenced. Primers are listed in Table 2.1.

2.13.1.4 Restriction enzyme digest of isolated plasmids

A restriction enzyme digest was performed as another way of determining that the plasmid had contained the right insert. The restriction enzyme digest protocol involves several restriction enzymes which cuts the plasmid at specific sites in order to identify if the plasmid purchased was the correct one.

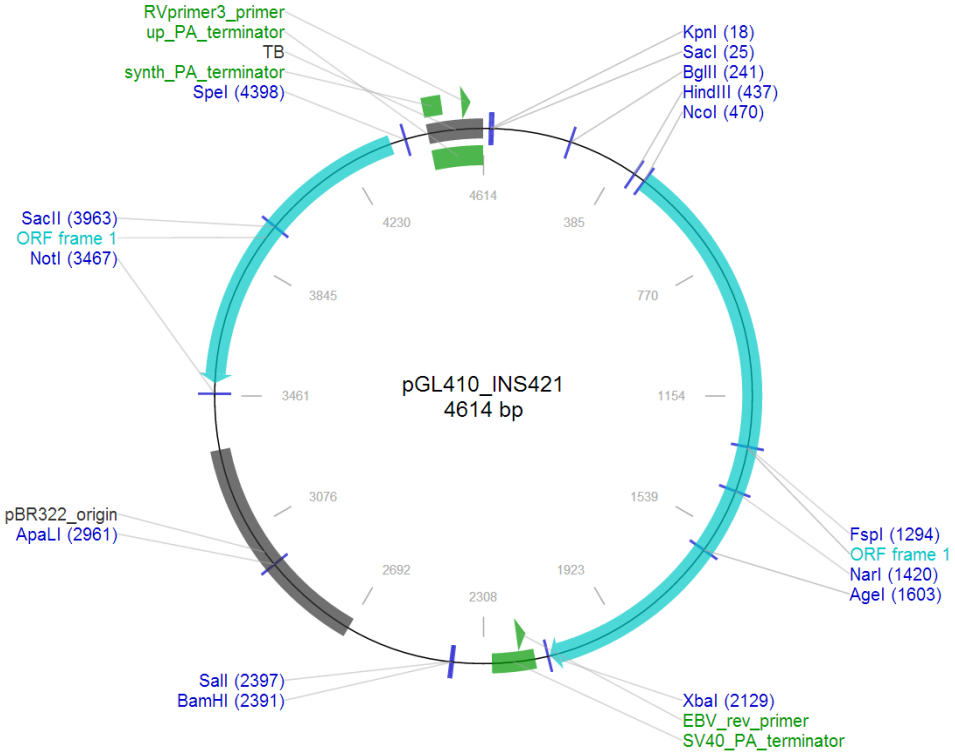
Two reactions for each plasmid were performed to check the identity of the plasmid. In a total reaction volume of 10 µL, one reaction consisted of 0.5 µL BglII, 0.5 µL BamHI, 1 µL 10x Buffer, 3 µL water and 5 µL plasmid. The other reaction consisted of 0.5 µL NotI, 0.5 µL BamHI, 1 µL 10x Buffer, 3 µL water and 5 µL plasmid. The tubes were incubated at 37°C for two hours and then visualised on a 3% agarose gel to determine the lengths of the cut plasmid (Figure 2.9B). Alongside the PCR products, a 1kb ladder was also loaded.

2.13.1.5 Sodium bisulfite treatment of the plasmids

To detect methylation at the CpG sites in the insulin promoter, the plasmids had to be bisulfite converted. This was performed using the EZ DNA methylation kit from Zymo according to manufacturer's instructions. All centrifugation steps were performed at 14,000 rpm for 30 seconds.

In each reaction, 5 µL of M-Dilution Buffer was added to 500 ng of plasmid DNA. The samples were adjusted to 50 µL with water. The samples were incubated at 37°C for 15 minutes. After the incubation, 100 µL of the CT Conversion Reagent was added to each sample and then was left to incubate at 50°C for 16 hours. After this step, the samples were incubated on ice for 10 minutes and 400 µL of M-Binding Buffer was added to the samples. This buffer helps the samples to bind to the column. The samples were loaded onto a spin column placed into a 2 mL collection tube and centrifuged. The flow through was discarded and 200 µL of M-Wash Buffer was added to the columns and spun. 200 µL of M-Desulphonation Buffer was added to the columns and was left to incubate at RT for 15 minutes. Unmethylated cytosines would then convert to uracil upon desulphonation. After incubation, the columns were centrifuged. The flow through was discarded and 200 µL of M-Wash Buffer was added to the column and centrifuged. This step was repeated. The columns were placed into new 1.5 mL collection tubes and 10 µL M-Elution Buffer was added to the column, directly onto the column matrix. The columns were spun to elute the bisulfite converted DNA.

A



B

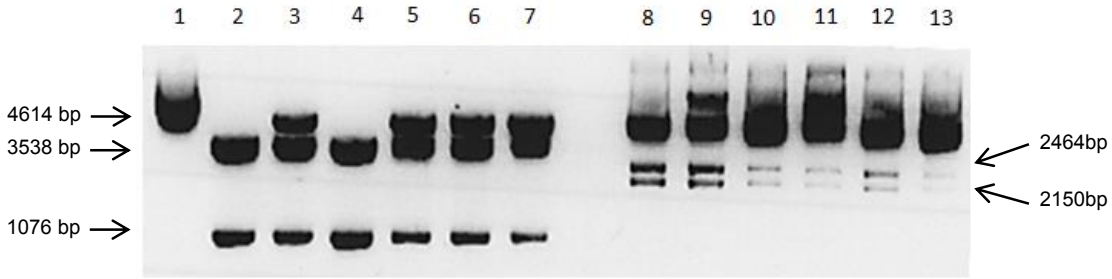


Figure 2.9. Restriction enzyme digest of plasmid DNA. The insulin promoter was inserted into the pGL410 plasmid and deposited into Addgene by the Ferreri lab (Kuroda *et al.*, 2009). Taken from http://www.addgene.org/49057/ (A). A restriction enzyme digestion was performed to quickly determine the identity of the plasmid. Three restriction enzymes were used in two separate reactions (B). The products were run on a 3% agarose gel with a 1kb ladder. The first reaction consisted of BamHI and NotI, resulting in products of 3538 bp and 1076 (Lanes 2-7). The second reaction consisted of BamHI and BglII, resulting in products of 2464 bp and 2150 bp (lanes 8-13). The uncut plasmid was loaded into Lane 1.

2.13.2 Methylation-specific PCR

Methylation specific PCR (MSP) was developed in 1996 (Herman *et al.*, 1996) and is highly sensitive in detecting methylated or unmethylated changes in DNA in a particular region. MSP is based on quantitative real-time PCR (qPCR) whereby a targeted DNA molecule is amplified and quantified. In qPCR, detection chemistries allow the investigator to visualise real time amplification during the early phase of the reaction. In this thesis, primer sets that were specific to unmethylated and methylated CpG sites were designed. Under optimal conditions, there was an initial hot start to complement the taq polymerase, followed by a denaturation step, whereby the double stranded DNA is separated in order for the polymerase to help assemble a new DNA strand. The annealing step involves the primers binding to the single stranded DNA which then leads to the elongation step. These steps are repeated over many cycles, normally 35-45. These reactions were quantified and recorded in real time, producing amplification curves. Two common detection chemistries used in qPCR are SYBR Green and Taqman. Both report amplified DNA in different ways.

2.13.2.1 SYBR Green

SYBR Green is a dye that binds to double-stranded DNA. As SYBR Green intercalates to any double-stranded DNA, it results in non-specific detection of nucleic acids (Zipper *et al.*, 2004). After each cycle, once the dye binds to the DNA, it fluoresces and it is picked up by the thermal cycler (Figure 2.10). Normally after a run with SYBR Green, a melt curve is performed. A melt curve checks for the specificity of the assay by melting the amplicons. If unspecific binding or primer dimers occur, it is easily detected by the different peaks on the meltcurve as every product has a different dissociation temperature. For example, primer dimers tend to dissociate at a lower temperature than the amplicon of interest. Aside from this, is it advantageous to use SYBR Green as it can be performed at low cost and the experimental design is relatively straight-forward.

2.13.2.2 Taqman

Taqman technology incorporates taq polymerase, primers but also has the addition of a hydrolysis probe. This probe is an oligonucleotide that has a fluorescent reporter dye attached to the 5' end and a quencher on the 3' end (Figure 2.11). The

reporter dye is excited by the machine and its energy is passed on to the quencher by a process known as Fluorescence Resonance Energy Transfer (FRET) (Holland *et al.*, 1991). As the probe binds to the amplicon during each cycle, it is intact and the quencher reduces the fluorescence emitted by the reporter dye, therefore no fluorescence is detected by the machine. However, when the Taq polymerase extends from the primer, it cleaves the probe due to the 5'-3' exonuclease activity. This separates the reporter dye from the quencher, in turn increasing the fluorescence of the reporter dye. As this method is probe-based, the detection is more specific than non-probe based methods such as SYBR Green, which binds to any double-stranded DNA. Although the Taqman system is more costly than SYBR Green, it is more specific and has multiplex capabilities.

2.13.2.3 Primer and probe design

Primer design for DNA methylation is slightly different to standard qPCR but both must be optimized in order to perform a highly efficient assay. The main difference between designing primers for DNA methylation and standard qPCR, is that the change in DNA must be considered after bisulfite conversion. All non-CpGs cytosines and unmethylated cytosines at CpG sites will have converted to uracils. There are guidelines for optimal primer design for methylation assays (Li and Dahiya, 2002). For bisulfite-specific PCR (BSP), this guideline suggests that primers must not contain CpG sites to avoid discrimination against methylated or unmethylated DNA. For MSP, primers should have at least one CpG site and methylated and unmethylated primer pairs should have the same CpG sites. Generally, the length of primers should ideally be between 26-32bp, have similar annealing temperatures and the amplicon should ideally be between 100-300bp long. For Taqman assays, an additional probe has to be designed. The melting temperature (Tm) of the probe must be 8-10°C higher than the primers and must be placed near the primers.

For the SYBR Green assay, there were three primers sets (Table 2.1). The BSP primer set did not discriminate between methylated and unmethylated CpG sites. The MSP1 primer set was positioned to detect unmethylated CpG sites at -357 and -69. The MSP2 primer set was positioned to detect unmethylated CpG sites at -206 and -135 (Figure 2.12). For the Taqman assay, there were two primer sets, BSP2 and MSP2 (Table 2.1).

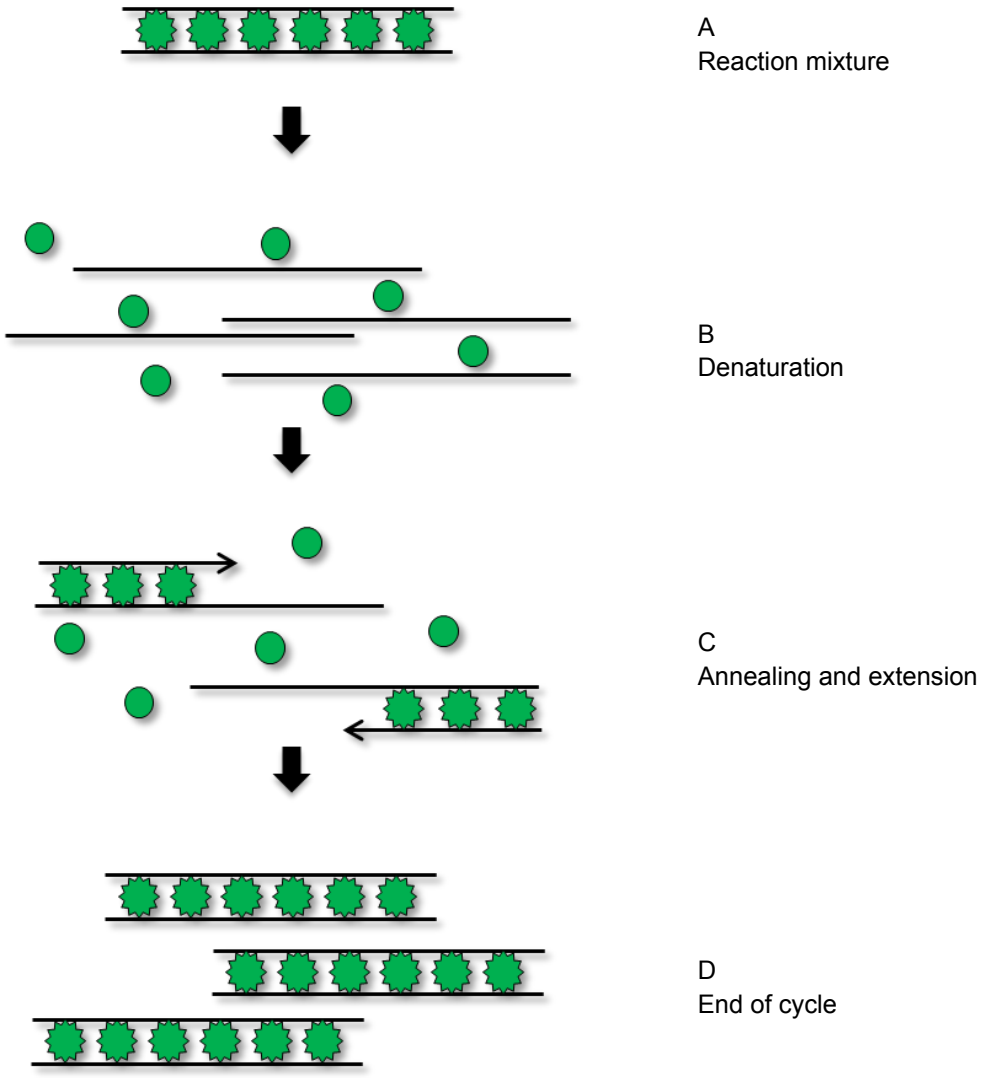


Figure 2.10. SYBR Green chemistry mechanisms in qPCR. In a qPCR reaction, the SYBR Green dye fluoresces when bound to double-stranded DNA (A). At the start of the reaction, DNA was denatured, releasing the SYBR Green dye. This leads to the reduction of fluorescence (B). During the annealing and extension step, primers anneal to the template strand and then are extended by the polymerase and free nucleotides (C). Once the PCR products are generated, the SYBR Green dye binds to the double-stranded product, increasing the fluorescence that is detected by the ABI 7500 Real Time PCR System (D).

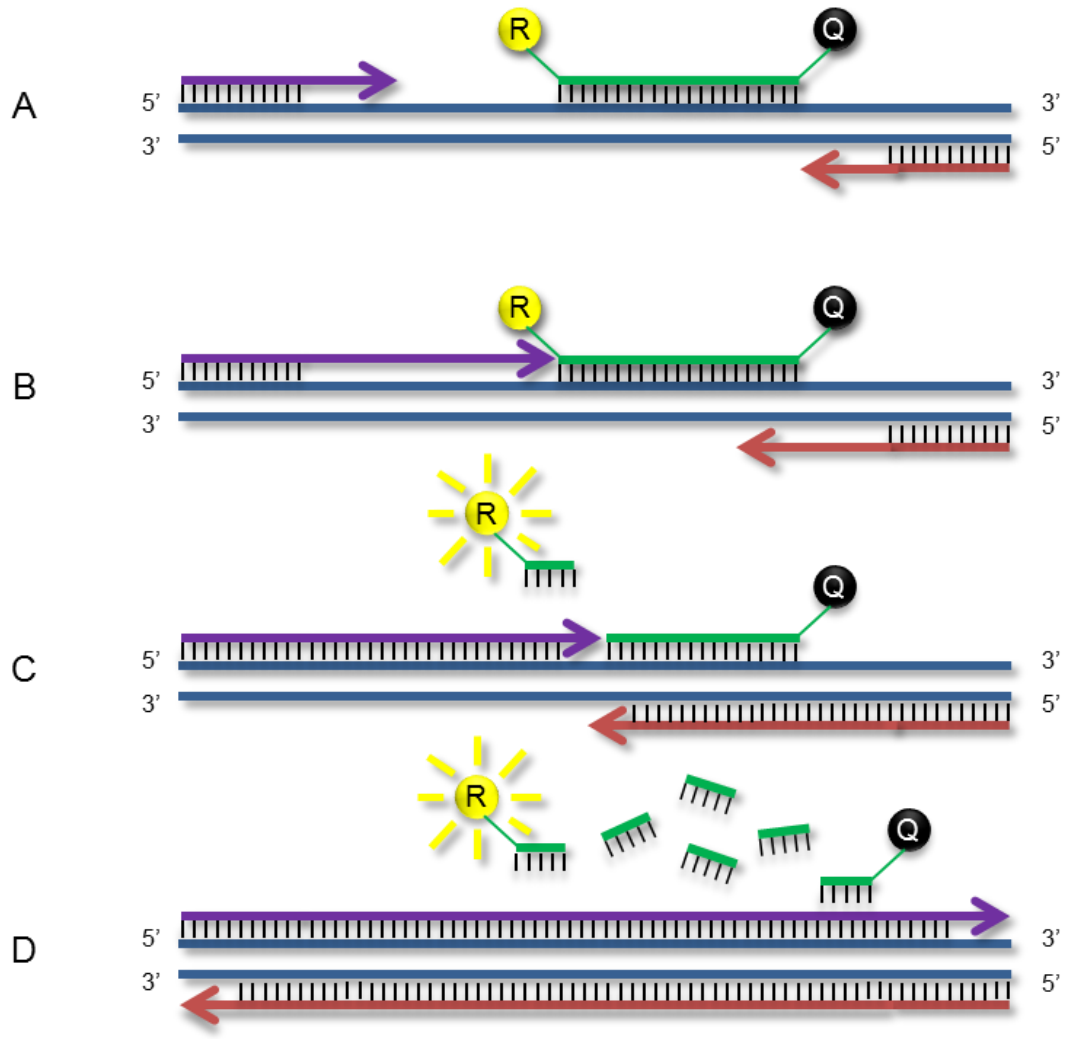


Figure 2.11. Taqman probe chemistry mechanisms in qPCR. At the start of the reaction, the probe and primers annealed to the DNA strand. Whilst the probe was intact, the reporter dye and quencher were in closer proximity to each other, allowing FRET to occur (A). The DNA polymerase then starts to extend the primer with dNTPs (B). Once the polymerase reaches the probe, the 5' nuclease activity of the polymerase, cleaves the probe (C). Once the probe was cleaved, the released reporter molecule will no longer be quenched (D).

Table 2.1. Primer sets designed for detecting methylation differences in serum.

Primer	Primer Sequence	Bases	Product size (bp)	Tm
RVpri3	CTAGCAAAATAGGCTGTCCC	20		60.7
MSP1	F: TGGGGATAGGGGTTTGGGGATAGTAGT	27	350	71.6
	R: CTCCCCTACCTCTCAACCCCTACCA	25		71.5
MSP2	F: TGGGTTTTTGGTTAAGATTTTAATGATTT	29	130	65.6
	R: CAACAAATAACTAAAAACTAAAACTACAATTTCCA	35		65.2
BSP	F: TGGGGATAGGGGTTTGGGGATAGTA	25	350	71.0
	R: CATCTCCCCTACCTCTCAACCCCTAC	26		70.1
BSP2	F: TTTAGTTGTGAGTAGGGATAGGTTTGGTTA	30	170	66.0
	R: TCCCCTACCTCTCAACCCCTAC	22		66.2
Probe	[6FAM]TGGTTTTGAGGAAGAGGTGTTGATGATTAAGGA[BHQ1]	33		74.1

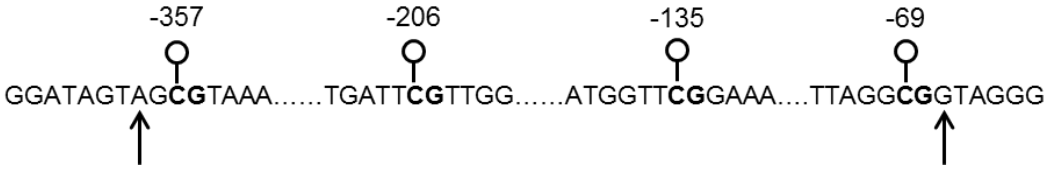


Figure 2.12. Primer design and selection. MSP1 targeted the CpG sites at -357 and -69 upstream of the transcription start site. MSP2 targeted the CpG sites -206 and -135 upstream of the transcription start site. The 3' ends of the BSP primers sets were placed just upstream and downstream of the MSP1 primer set (black arrows). The BSP primers did not discriminate between methylated and unmethylated CpG sites therefore providing the total amplifiable region for nested-MSP. The 3' ends of the BSP2 primer sets were placed upstream and downstream of the MSP2 primer set. The probe was placed in between the MSP2 primers.

2.13.3 Amplification efficiency

High amplification efficiency is needed to ensure that with each cycle, the PCR product is amplified correctly and consistency. For example, the optimum reaction efficiency should be 100%, this is where PCR products doubles exactly in each cycle. The plasmids that were purified from Section 2.13.1.2 were used to create a 10-fold serial dilution series in order to test the amplification efficiency of the assay. Selection of the primers targeting specific CpG sites was based on available literature (Husseiny *et al.*, 2014). The qPCR reaction mixture of 10 µL in total consisted of 5 µL SYBR Green, 200 nM forward and reverse primers, 2 µL bisulfite converted genomic DNA and the rest was made up of water. Amplification and melt curves were analysed using the ABI Real-Time 7500 software. The products were run on a 3% agarose gel.

A Taqman assay was also performed along with the SYBR Green assay to see if the increased specificity of Taqman, would result in a highly efficient assay. For the Taqman assay, a total reaction volume of 25 µL consisted of 12.5 µL Taqman master mix, 200 nM probe, 400 nM primers, 2 µL of bisulfite converted genomic DNA template and the rest was made up with water. Amplification curves were analysed using the ABI Real-Time 7500 software and the products were run on a 3% agarose gel. Efficiency was calculated using the quantification cycle (Cq) values. A Cq value is the number of cycles at which the fluorescent signal of the reaction crosses the threshold. Cq is used to calculate DNA copy number as the Cq value is inversely related to the starting amount of target i.e. a lower Cq correlates with higher target expression. The Cq values from a dilution series are plotted as a standard curve where the slope of the straight line and its intercept was used to calculate the efficiency:

$$\text{Efficiency (\%)} = (10^{(-1/\text{slope})}-1)*100$$

Where the slope is -3.32, this would result in 100% efficiency. This means that for every cycle, the PCR products double exactly. Amplification efficiency was carried out three times. The primer efficiencies with SYBR Green were for BSP was 98.5±3.4%, MSP1 98.7±4.1% and MSP2 90.5±6.6%. Primer efficiencies with Taqman were: BSP2 79% and MSP2 88%.

2.13.4 Distinguishing between methylated/unmethylated DNA

To ensure that the primers were specific to unmethylated CpGs, an aliquot of the plasmid was fully methylated in a reaction with M.SssI CpG methyltransferase, which in turn, methylates all cytosine resides.

In a 0.2 mL PCR tube with the total reaction volume of 20 µL, 1 µg plasmid DNA was added to 12 µL water, 2 µL 10x Buffer 2, 2 µL of diluted SAM and 1 µL of M.SssI. The reaction was incubated at 37°C for two hours in a thermal cycler. The reaction was stopped by heating the reactions at 65°C for 20 minutes. The qPCR reaction mixture of total 10 µL, consisted of 5 µL SYBR Green, 200 nM forward and reverse primers, 2 µL bisulfite converted genomic DNA and the rest was made up of water. Amplification and melt curves were analysed using the ABI Real-Time 7500 software and the products were run on a 3% agarose gel.

2.13.5 Nested methylation-specific PCR

Nested-MSP involved two rounds of amplification, normally with PCR-product from the first qPCR reaction being used as a template for the second round of qPCR. This is especially useful if there were low concentrations of starting material or to increase the specificity of the assay. In this method, nested-MSP was required as the starting material was DNA from serum, hence low DNA concentrations. Also, this assay needed high specificity to detect single base changes in the DNA (Figure 2.13).

2.13.5.1 First step MSP

For all nested-MSP assays, the serum samples were prepared with DNA extraction and bisulfite conversion as described in Section 2.7 and 2.13.1.15 respectively. The only modification to the DNA extraction step was that the samples were eluted in 30 µL AE Buffer instead of 150 µL. Nested-MSP was performed in a 96-well plate with a total reaction volume of 10 µL in duplicate. The primer sets used were BSP and MSP1. The master mix prepared for 1 reaction contained 5 µL QuantiTect® SYBR® Green, 0.2 µL forward sequence primer, 0.2 µL reverse sequence primer and 2.6 µL sterile water. MSP was performed on an ABI 7500 machine (Applied Biosystems, UK) with the first step undergoing an initial denaturation and enzyme activation step at 95°C for 15 minutes, followed by 15 amplification cycles of 95°C for 15 seconds and an annealing and elongation step at 60°C for one minute. A melting curve

analysis was performed afterwards. A reaction with the plasmid instead of the patient sample was used as a positive control. A no-template control (NTC) was also included.

2.13.5.2 Clean up of PCR products

After the first step, the PCR products were cleaned up using a Qiagen PCR Purification kit following the manufacturer's instructions. Cleaning up the reaction removes primers and salts from the previous reaction ensuring that it does not inhibit or interfere with the second step. All centrifugation steps were performed at 14,000 rpm at RT unless specified otherwise.

The PCR products were transferred to 1.5 mL eppendorf tubes and five volumes of Buffer PB was added to one volume of the PCR samples. The samples were mixed thoroughly and applied to the QIAquick column and centrifuged for one minute. The flow-through was discarded and the column was placed back into the same collection tube. 750 µL Buffer PE was added to the column and centrifuged for one minute. The flow-through was discarded, column placed back in the same collection tube and centrifuged the column for an additional minute to remove residual buffer. The column was placed in a new 1.5 mL microcentrifuge tube and 30 µL Buffer EB was added to the column. The column was left to stand at RT for one minute before spinning for one minute.

2.13.5.3 Second step MSP

The second step MSP was also performed on the ABI 7500 machine. The master mix prepared for one reaction contained 5 µL QuantiTect® SYBR® Green, 0.2 µL forward sequence primer, 0.2 µL reverse sequence primer, 2.6 µL sterile water. The MSP conditions were as follows with primer sets BSP and MSP2: an initial denaturation and enzyme activation step at 95°C for 15 minutes, followed by 40 amplification cycles of 95°C for 15 seconds and annealing and elongation step at 57°C for one minute. A melting curve analysis was performed after. A reaction with the plasmid instead of the patient sample was used as a positive control. A NTC was also included.

2.13.6 Optimisation of nested-MSP assay

Initially, DNA methylation was studied in exon 2 of the insulin gene as CpG sites in that region were shown to be unmethylated in a tissue-specific manner (Akirav *et al.*, 2011). The primer sets used were against methylated and unmethylated CpG sites (Appendix I, 8.1.8). DNA extraction and bisulfite conversion was performed as described in Section 2.7 and 2.13.1.5 respectively. Nested-MSP was carried out involving standard PCR in a thermal cycler first, and then qPCR for the second step using SYBR Green. However, there was non-specificity between the two methylated and unmethylated primers.

The next step was to focus on the promoter region, as this region was reported to also have unmethylated CpG sites (Kuroda *et al.*, 2009). Here, both steps in the nested-MSP were performed with qPCR and SYBR Green. Optimisation steps included performing a primer matrix, using varying primer concentrations, designing new shorter primer sets, the use of different SYBR Green master mixes, use of different machines and including dimethyl sulfoxide (DMSO) in the reactions. After the optimisation steps, primer dimers appeared along with the products of interest. Therefore, Taqman was then used as it was more specific and it would be less likely for primer dimers to appear.

2.13.7 Taqman assay to assess methylation

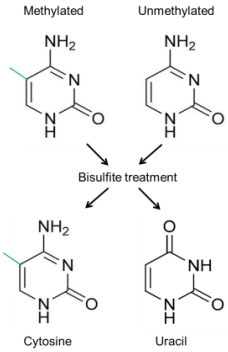
As in Section 2.13.5.1, the samples were prepared with DNA extraction and sodium bisulfite. Unlike the nested-MSP, Taqman was carried out as a one-step assay as with the two-step assay, no PCR products were amplified. qPCR was performed in a 96-well plate with a total reaction volume of 25 µL. The primer sets used were BSP2 and MSP2. The master mix prepared for one reaction contained 12.5 µL Taqman gene Expession, 400 nM forward and reverse primers, 200 nM probe, 2.6 µL sterile water and 2 µL of the bisulfite converted template. qPCR was performed on an ABI 7500 machine with the following conditions: 50°C for 2 minutes, initial denaturation and enzyme activation step at 95°C for 10 minutes, followed by 40 amplification cycles of 95°C for 15 seconds and an annealing and elongation step at 60°C for one minute. A reaction with the plasmid instead of the patient sample was used as a positive control. A NTC was also included.

2.13.8 Sequencing of PCR products

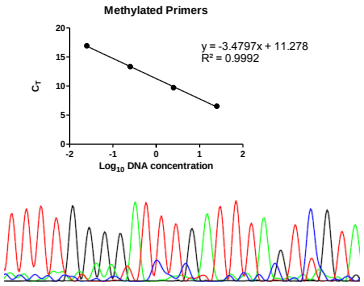
After running the PCR products on a 3% agarose gel, the bands were cut out and cleaned up using the QIAquick Gel Extraction Kit according to manufacturer's instructions. All centrifugation steps were performed at 14,000 rpm at RT unless specified otherwise.

Six volumes of Buffer QG were added to one volume of gel in a 1.5 mL eppendorf tube. The tubes were incubated for 10 minutes until the gel slice was completely dissolved. One volume of isopropanol was added to the sample, applied to a spin column and centrifuged for one minute. The flow-through was discarded and the column was placed back into the same collection tube. 500 µL Buffer QG was added to the column and centrifuged for one minute. The flow-through was discarded and the column was centrifuged for an additional minute. The column was placed in a new 1.5 mL tube, 30 µl of Buffer EB was added to elute the DNA and spun for one minute. An aliquot of the eluted sample was sequenced (Figure 2.14).

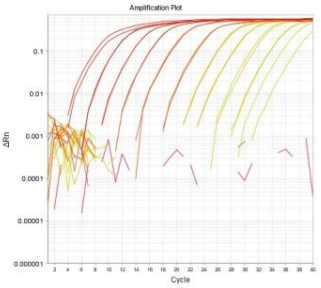
A
Growth of bacteria with transformed plasmid.
Plasmid DNA extraction.



B
Bisulfite conversion of plasmids and samples.



C
Amplification efficiency and sequencing.



D
Nested-MSP.

Figure 2.13. Summary of the protocol for monitoring β-cell death in serum samples. The transformed plasmid with the insulin promoter insert was grown, isolated and purified using the GeneJET Plasmid Miniprep kit (A). DNA from serum was extracted and treated with sodium bisulfite. Once the integrity of the plasmid was assessed on an agarose gel, the plasmid and patient samples were treated with sodium bisulfite (B). Amplification efficiency was performed and the PCR products were also sequenced (C). Patient samples were analysed using nested-MSP (D).

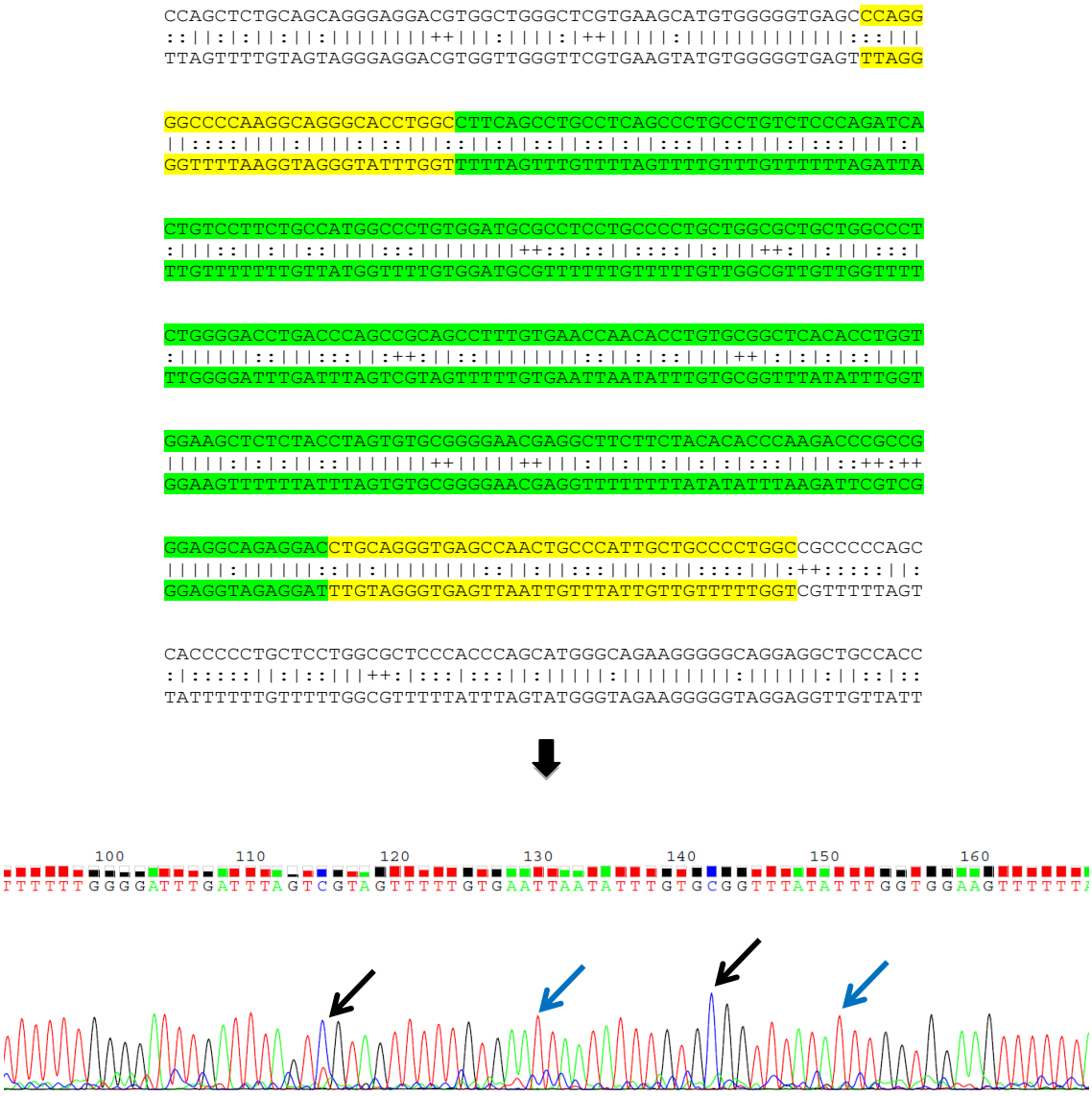


Figure 2.14. An example of sequencing results of the insulin promoter region. The insulin promoter region was sequenced to ensure the primers were amplifying the correct region. At first, the original DNA sequence (top line) was treated with sodium bisulfite to convert unmethylated cytosines into uracils, in which after amplification would turn into thymines (bottom line). Once sequenced, methylated CpGs were distinguished from unmethylated CpGs from the chromatogram. Methylated cytosines remained protected (black arrows) whereas unmethylated cytosines were converted (blue arrows).

2.14 Detection of β-cell death using pyrosequencing

Another technique to analyse DNA methylation is pyrosequencing. Pyrosequencing is a sequencing-by-synthesis method, in which during DNA elongation, pyrophosphate is released and then detected bioluminometrically (Elahi and Ronaghi, 2004). The assay involves several main steps (Figure 2.15). Bisulfite converted DNA was amplified using forward and reverse primers, one of which was biotinylated. The PCR products were then converted into single-stranded DNA (ssDNA) and retained by the biotinylated primer, as a template for the pyrosequencing reaction. A separate sequencing primer anneals to the DNA. When the DNA strand elongates, upon incorporation of a nucleotide, a pyrophosphate molecule is released and ATP-sulfurylase converts it to ATP, providing the energy for luciferase to generate visible light peaks that can be recorded using a camera. Pyrosequencing can be used as a validation method once CpG sites have been identified in methylation studies.

Genomic DNA was extracted from patient serum and treated with sodium bisulfite as previously described in Section 2.13.7. The pyrosequencing assay was designed to investigate the methylation status of two CpG sites that were previously studied in Section 2.13.5 and 2.13.7. PCR and sequencing primers were designed with the PyroMark Assay software (Qiagen). For PCR, the total length of the amplicon was approximately 200 bp. The total reaction mixture of 60 µL, consisted of 12 µL 5x Buffer, 0.4 µL DNA polymerase and 300 nM of the following primers: forward 5'-GGTTTGGTTATAGGGTTTTTGGTTAAGA-3' and reverse biotinylated 5'-[Btn]CCCATCTCCCCTACCTCTCA-3'. The rest was adjusted with water. The PCR thermal cycling conditions were as follows: denaturation at 95°C for one minute; 40 cycles of 95°C for 15 seconds, 55°C for 15 seconds, and 72°C for 10 seconds. PCR reactions were carried out with 3 µL of bisulfite-converted genomic DNA. A biotin-labelled primer (reverse primer) was used to purify the PCR products with streptavidin-coated Sepharose beads. PCR products were bound to the Sepharose beads and washed, denatured by using NaOH solution, and washed again. Then pyrosequencing primer 1 (5'-GTTATAGGGTTTTTGGTTAAGAT-3') was annealed to the purified single-stranded PCR product and the pyrosequencing was performed on a PyroMark Q96 ID (Qiagen) according to the manufacturer's instructions. The first pyrosequencing primers were stripped off and the second pyrosequencing primer was added (5'-TTTTATAGATTTAGTATTAGGGAAA-3'). Two sequencing primers were used as the CpG sites were too far apart for a single sequence to analyse.

2.14.1 Statistical analysis

Data analysis was performed using PyroMark Q96 from Qiagen. Methylation levels from each twin pair were analysed using paired and unpaired t-tests using SPSS Statistics v17.0, using $p < 0.05$ to be considered statistically significant.

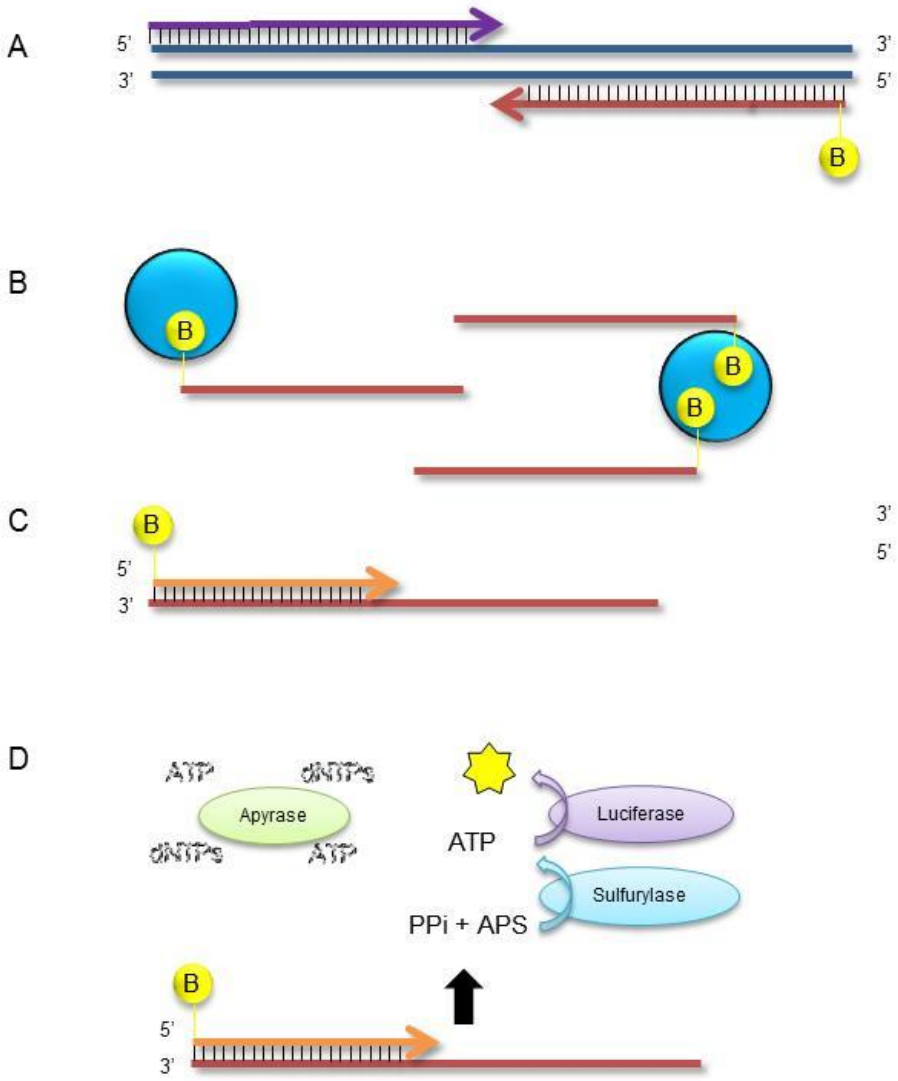


Figure 2.15. Summary of the pyrosequencing process. A PCR reaction involves two primers (forward and reverse) one of which is biotinylated (A). Once amplification is completed, the biotinylated amplicons were captured by binding to streptavidin-coated Sepharose beads (B). The DNA was denatured to produce ssDNA templates for the pyrosequencing assay (C). The ssDNA is released and the sequencing primer anneals to the template. During the pyrosequencing reaction, the primer is extended with the nucleotides (D). The incorporation of each nucleotide releases a pyrophosphate (PPi) molecule. Together with adenosine 5' phosphosulfate (APS), adenosine triphosphate (ATP) sulfurylase converts PPi to ATP. ATP provides the necessary energy to drive the luciferase conversion of luciferin to oxyluciferin. This in turn emits light that was detected by a camera. The enzyme apyrase degrades ATP and any unincorporated dNTPs.

Chapter 3

Analysis of DNA methylation in immune cells in type 1 diabetes

3 Analysis of DNA methylation in immune cells associated with type 1 diabetes

3.1 Introduction

This chapter focuses on the role of DNA methylation in T1D in immune cells. This study was carried out as part of the BLUEPRINT European programme where it received funding for €30 million (Adams *et al.*, 2012). BLUEPRINT is part of a larger consortium, the International Human Epigenome Consortium (IHEC), involving 39 leading European universities, research institutes and small companies. The main aim of BLUEPRINT is to provide at least 100 reference epigenomes and to investigate epigenetics in blood-based diseases such as leukaemia and T1D by focussing on cell types of the haematopoietic system. Identification of potential epigenetic markers may eventually lead to the development of novel and more individualised medical treatments.

T1D is an autoimmune disease characterised by the loss of insulin secreting β-cells, which in turn leads to exogenous insulin being injected into the patient (American Diabetes, 2004). Several genes such as *PTPN22*, *INS* and *HLA* confer susceptibility to T1D (Howson *et al.*, 2011). As the development of T1D involves the interplay between genetic and non-genetic factors, epigenetics have been suggested to contribute to the non-genetic part of T1D aetiology. DNA methylation is the best characterised epigenetic mark, particularly in diseases such as SLE (Absher *et al.*, 2013) and cancer (Irizarry *et al.*, 2009). Performing EWAS in MZ twins is useful as it removes genetic confounding factors. Recently, studies have identified several genes that were epigenetically modified in T1D. These genes include *HLA-DQB1*, *NFKB1A*, *GAD2* (Rakyan *et al.*, 2011a) *INS*, *IL-2RB* (Stefan *et al.*, 2013) and *IL2RA* (Belot *et al.*, 2013).

DNA methylation was studied in T1D as one of the cell types implicated in disease aetiology, $CD14^{+}$ monocytes, was profiled in a successful pilot study identifying several T1D-MVPs (Rakyan *et al.*, 2011a). The methylation status in monocytes was reported in 15 MZ twin pairs discordant for T1D. They showed that intra-pair DNA methylation differences at T1D-MVPs ranged from 0.13% to 6.6%. Any T1D-MVPS identified in this present study, can be compared to the T1D-MVPs calls made in the preliminary study to see whether the methylation marks are stable in the same twin pairs. Also, any T1D-MVPs identified in BLUEPRINT can be compared to healthy

reference epigenomes that other BLUEPRINT partners are investigating. To investigate DNA methylation changes in T1D, blood and buccal samples from 24 pairs of MZ twin pairs discordant for T1D were taken. Studying different cells types as opposed to PBMCs is important as each cell displays its own methylation profile (Reinius *et al.*, 2012). Four cell types were isolated from peripheral blood from heparin tubes. $CD14^{+}CD16^{-}$ monocytes were chosen as it has been reported to show abnormal expression in type 1 diabetic patients (Beyan *et al.*, 2010; Padmos *et al.*, 2008). In T1D, the pancreas is infiltrated with autoreactive T cells leading to insulitis and β-cell destruction (Roep, 2003), hence studying the $CD4^{+}$ cell type. $CD19^{+}$ B cells have been reported to be associated with T1D and undergo T-cell independent antigen activation (Pescovitz *et al.*, 2009). Buccal samples provide another target tissue which is as easily obtainable as peripheral blood. In a recent study, it was shown that a buccal sample revealed more than six times hypomethylated regions than blood (Lowe *et al.*, 2013). Therefore buccal samples may be useful in identifying new regions of methylation differences that may not be picked up in blood.

In this study, different DNA methylation profiling technologies were used. BS-seq requires bisulfite converted DNA to assess whole genome-wide methylation status. It remains the 'gold' standard method, however it is costly. Illumina450K was used to interrogate over 485,000 CpG sites and comprises of a different technology compared to the Illumina27K, which was used in the preliminary study.

3.2 Aims and objectives

The aim of this study was to identify any differences in DNA methylation in MZ twins discordant for T1D. PBMCs were isolated from whole blood using Percoll. The different cell types were sorted by MACS and then assessed for purity by FACS. DNA was extracted from the cells and treated with sodium bisulfite. The samples were then hybridized onto the Illumina450K BeadChip and sequenced to profile DNA methylation in the different cell types. Bisulfite sequencing was also performed to generate a whole genome-wide DNA methylation profile as the Illumina450K does not cover all the CpG sites.

3.3 Results

3.3.1 Quality control of DNA methylation analysis

To identify potential DNA methylation differences in MZ twins discordant for T1D, a total of 24 twin pairs aged 43±17 years were ascertained and sampled. PBMCs were isolated from peripheral blood and sorted for $CD14^+CD16^-$, monocytes $CD4^+$ T cells and $CD19^+$ B cells using MACS. Buccal brush samples were also taken and 99 cord blood samples from new-borns dried onto card were also analysed. Cell purities determined by flow cytometry were 95±3% for monocytes, 97±2% for T cells and 83±12% for B cells. DNA from the sorted cells, buccal samples and cord blood samples were treated with sodium bisulfite and DNA methylation profiling was performed using Illumina450K.

The first batch of samples profiled from 16 twin pairs was analysed. The raw data were imported into R and analysed with the ChAMP pipeline (Morris *et al.*, 2014). Quality control and normalisation of the array were performed. To assess whether the samples from the twins were paired correctly, 65 SNPs were analysed in each sample and compared to the co-twin. SNPs were measured as in MZ twins, concordance of SNPs between a pair is 100%. When one twin was compared to the co-twin in the same pair, there was a high positive correlation between the twin pairs (Figure 3.1A). However, when the SNP probes from a diabetic twin were compared to another diabetic twin in another pair, there are few similarities (Figure 3.1B). This check ensured that there were no technical inaccuracies regarding twin sampling.

A principal component analysis (PCA) was performed to examine the differences or similarities between the different cell types. PCA is an algorithm for dimension reduction and identifies principal components, which are directions along which the variation in the data is maximal (Ringner, 2008). After performing PCA, it showed that DNA methylation patterns differed between cell populations but not differentiating individuals, for example, not differentiating between diabetic and non-diabetic twin (Figure 3.1C).

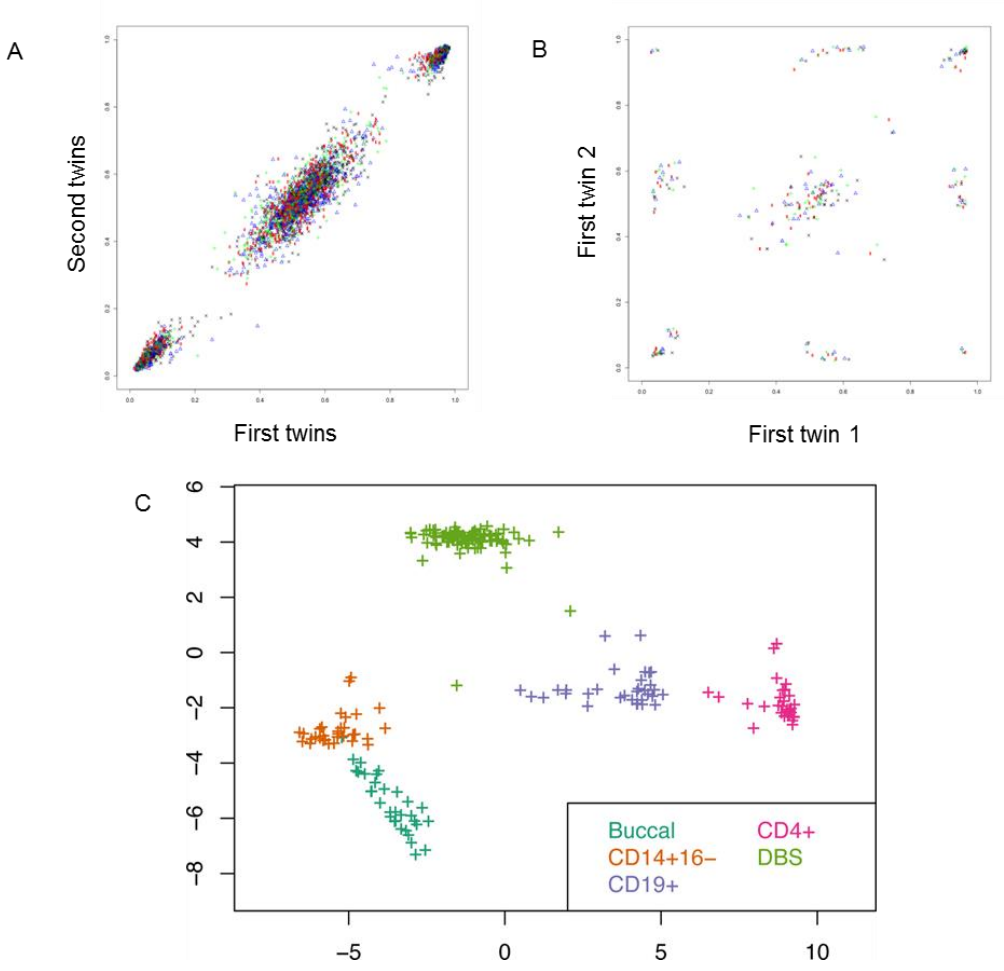


Figure 3.1. Quality control analyses of samples from 16 twin pairs. SNP analysis of diabetic vs. healthy twin, across all 16 pairs (n=65 SNP probes). Shown was a high correlation between identical pairs (A). As a negative control, 65 SNP probes were compared between unrelated individuals (B). Principal component analysis of the 1,000 most variable positions from the first batch of samples that included 16 twin pairs (C). Differences were seen between cell populations but no inter-individual differences.

3.3.2 DNA methylation signature is stable after three years

DNA methylation differences were found in $CD14^+$ monocytes in 15 MZ twin pairs in an earlier study using the Illumina InfiniumHumanMethylation27K BeadChip array (Illumina27K) (Rakyan *et al.*, 2011a), which assays over 27,000 CpG sites (Bibikova *et al.*, 2009). T1D-MVPs identified from the earlier study were compared to all four cell types in the present study (Figure 3.2). The correlation was most significant when comparing the same cell type, $CD14^+$ monocytes at 0.26, p-value=1.48e-03 (Figure 3.2A). The correlation was weaker when the comparison was made with other cell types, $CD19^+$ B cells correlation -0.17, p-value-9.71e-01 (Figure 3.2B), $CD4^+$ T cells correlation 0.10, p-value=1.30e-01 (Figure 3.2C) and buccal cells correlation 0.13, p-value=7.66e-02 (Figure 3.2D).

A comparison was also made between the twins from the preliminary study to different twin pairs to see whether DNA methylation differences would be detected in other twins. Of the 16 pairs of twins profiled, nine pairs were from the preliminary study. The correlation with the new individuals was -0.13, p=9.24e-01 (Figure 3.3A), whereas between the same sampled individuals, the correlation was stronger at 0.47, p=1.46e-08 (Figure 3.3B). Individual correlation plots for each of the 16 pairs is shown in Appendix III, Figure 8.5.

In summary, DNA methylation comparison between samples from the same pairs of twins taken years apart can be detected, suggesting that these marks are stable for at least three years.

3.3.3 Cord blood revealed weak correlation when compared to previously identified T1D-MVPs

In addition to the samples taken from the twins, 99 cord blood samples from a cohort of new-borns were processed. The new-borns were from a Swedish cohort, of which 50 went on to develop T1D and 49 that did not. Cord blood was studied to see if any significant methylation marks found in children or adults could be detected at birth. The cord blood samples were spotted onto card creating dried blood spots (DBSs) which were stored at 4°C until processing.

T1D-MVPs identified in $CD14^+$ from the preliminary study were compared to the DBS samples (Figure 3.3C). Differences found between the diabetic and non-diabetic DBS samples were compared to the T1D-MVPs from the preliminary study. The correlation was 0.20, p=1.25e-02 revealing a weak correlation. As cord blood from the new-borns was not isolated for different cell types, as was the case of the present study, the weak correlation suggests detection of any significant DNA methylation differences in unsorted samples is really low.

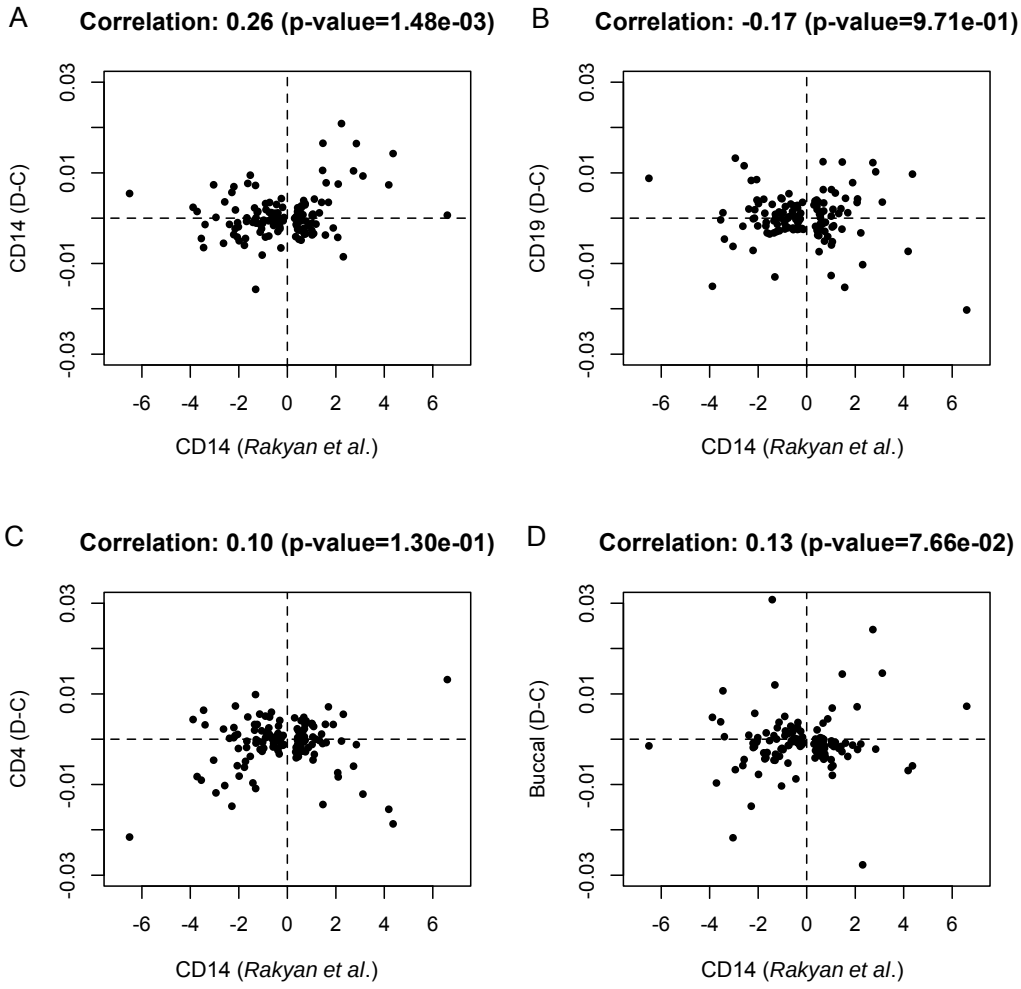


Figure 3.2. Correlation of MVP calls in T1D compared against previously identified T1D-MVPs. T1D-MVPs identified in the discovery cohort (vertical axis) plotted against previously identified T1D-MVPs from $CD14^+$ monocytes (horizontal axis) (Rakyan *et al.*, 2011a). Calls from Rakyan *et al*. were plotted against $CD14^+CD16^-$ monocytes (p = 1.48e-03) (A), $CD19^+$ B cells (p = 9.71e-01) (B), $CD4^+$ T cells (p = 1.30e-01) (C) and buccal cells (p = 7.66e-02) (D). The strongest correlation was observed between the monocytes. D: diabetic; C: control.

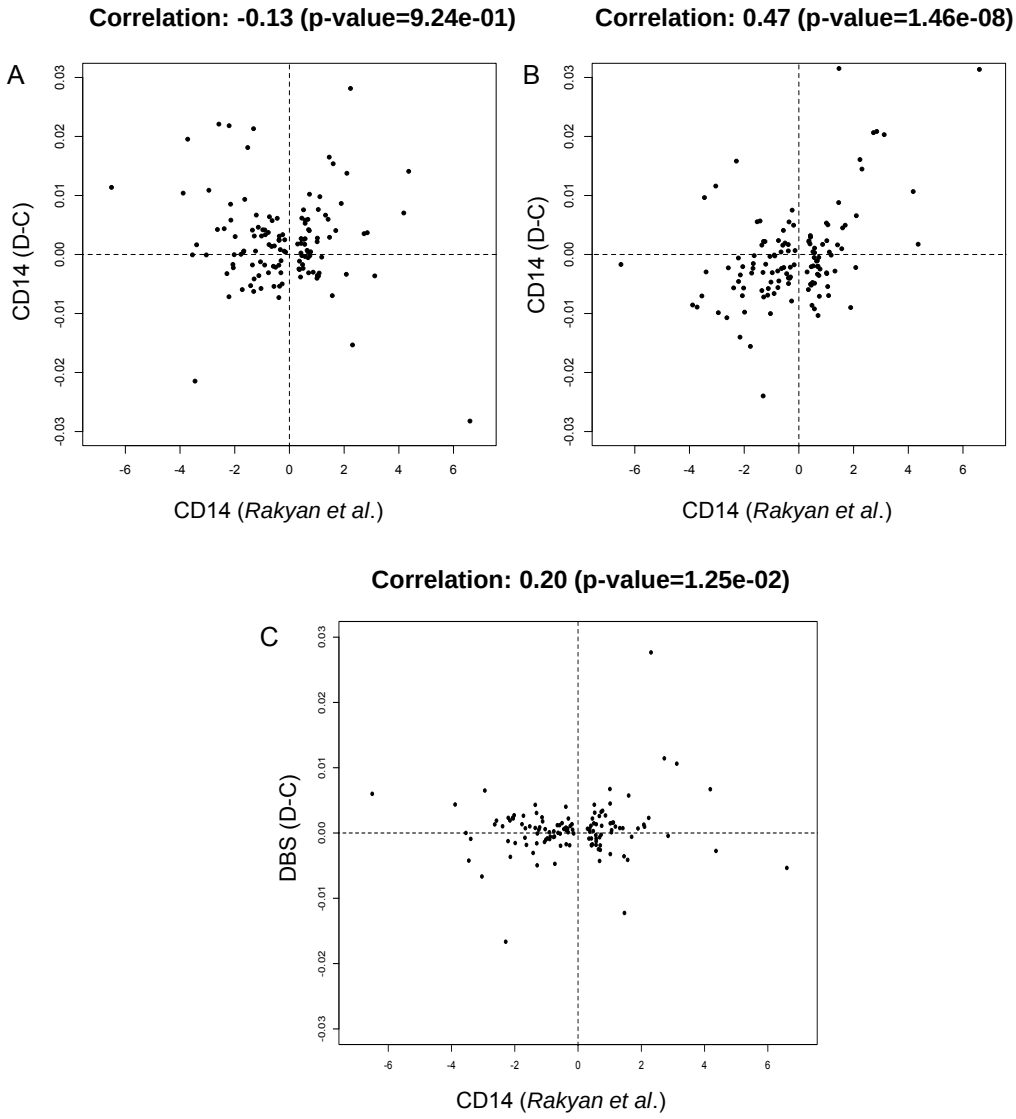


Figure 3.3. Correlation of MVPs calls in T1D compared against same twin pairs and cord blood. T1D-MVPs identified in the discovery cohort (vertical axis) plotted against previously identified T1D-MVPs from $CD14^{+}$ monocytes (horizontal axis) (Rakyan *et al.*, 2011a). Calls from Rakyan *et al*. were plotted against new twin pairs (p = 9.24e-01) (A), previously samples twin pairs (p = 1.46e-08) (B) and cord blood (p = 1.25e-02) (C). D: diabetic; C: control.

3.3.4 Discovery of DMPs using Illumina450K BeadChip platform

An additional eight pairs of twins were added to the array, therefore the overall analysis encompassed 24 twin pairs. As small methylation differences were observed, the question here was to see if there were any new and more significant DNA methylation differences between the 24 twin pairs in the four cell types. The previous analysis with 16 twin pairs revealed 356 DMPs in which they were cell-specific. $CD4^+$ T cells had the strongest signal with 166 DMPs, followed by $CD14^+CD16^-$ with 123, $CD19^+$ B cells with 45 and buccal cells with 22 DMP.

Quality control and normalisation of the array were performed resulting in the final probe set of 437,234 of a total 485,512. Hierarchical clustering showed clear separation of different cell types (Figure 3.4). The clustering was not based on disease status, age of sampling or age at diagnosis of the diabetic twin. The analysis also revealed 258 DMPs P<10-5, in which they were cell-specific (Figure 3.5). The $CD4^+$ T cells had the strongest signal with 170 DMPs, followed by the $CD19^+$ B cells with 79, $CD14^+CD16^-$ with 8 and buccal cells with 1 DMP.

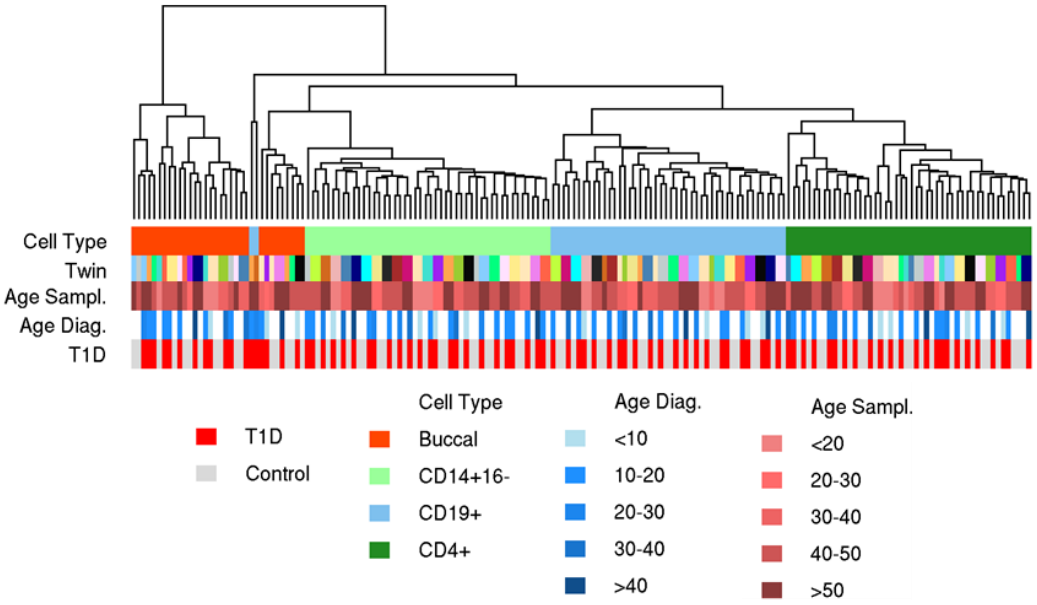


Figure 3.4. Hierarchical clustering of methylation similarities in different cell types. Methylation differences were seen between the four different cell types.

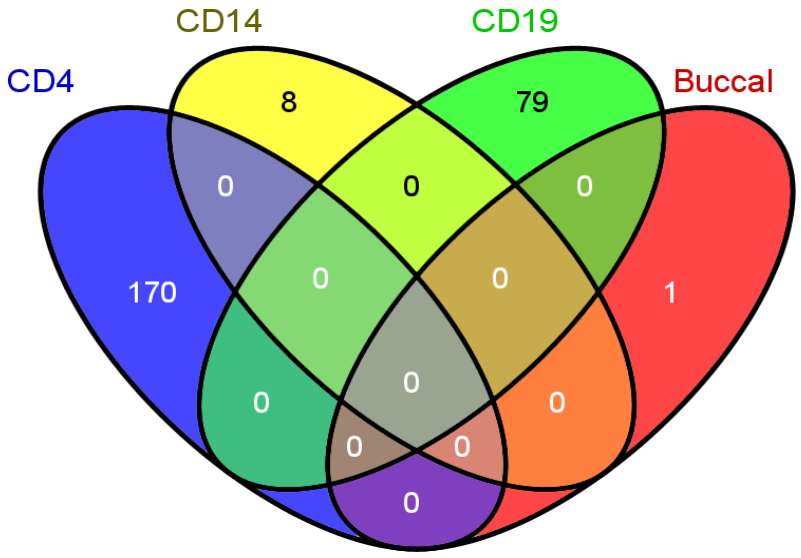


Figure 3.5. Cell type-specific DMPs. A total of 258 DMPs with p<10-5 were identified, all signals were cell type-specific, with $CD4^+$ cells obtaining the strongest signal.

3.3.5 Correcting for cell-type composition bias is important in EWAS

As the DMPs appeared to be cell type-specific, an analysis assessing cell composition of the different cell types was carried out (Figure 3.6). This was important as studying DNA methylation in whole blood can detect variations in DNA methylation between different cell types such as mononuclear cells and granulocytes (Reinius *et al.*, 2012). Since the purified cells were extracted to over 95% purity in the case with the monocytes and T cells and 85% for B cells, this minimised the epigenetic changes seen due to cell heterogeneity.

The proportions of the different cell types in the samples were determined (Figure 3.6A). In the $CD4^+$ samples, approximately 5% was estimated to have consisted of other cell types including B cells and $CD8^+$ T cells. In the $CD14^+CD16^-$ monocyte samples, approximately 5% was estimated to have consisted of $CD4^+$ and $CD8^+$ T cells and natural killer (NK) cells. In the $CD19^+$ samples, approximately 20% consisted of $CD8^+$ and $CD4^+$ T cells, $CD14^+$ monocytes, NK cells and granulocytes. In the cord blood samples, approximately 50% were granulocytes and the other 50% consisted of $CD8^+$ and $CD4^+$ T cells, NK cells, B cells and monocytes. The buccal samples also varied in cell composition. Approximately 40% were granulocytes, 25% were $CD4^+$ T cells and the rest were NK cells, B cells, monocytes but interestingly no $CD8^+$ T cells. A graph was constructed to show cell purity of each individual sample (Figure 3.6B). The spread of the $CD19^+$ purity analysis corresponds to the cell composition in Figure 3.6A.

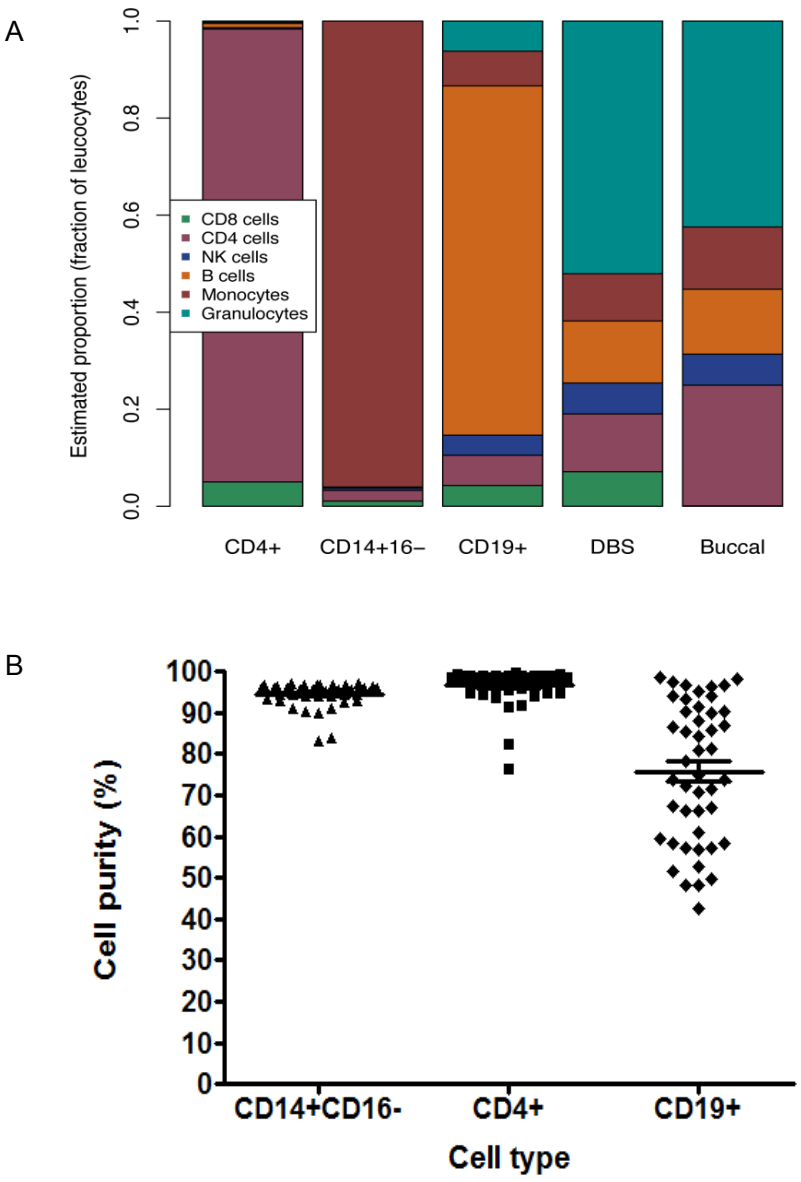


Figure 3.6. Analysing cell type composition in samples. The isolated cell populations were separated from PBMCs using Percoll. Cell purities were analysed using flow cytometry. The proportion of leukocytes in the samples was estimated using R (A). Individual purity values from each samples analysed using FACS were plotted (B).

3.3.6 T1D-MVPs are associated with genes involved in immunity

To assess the possible functional impact of the T1D-MVPs calls identified from these 24 twin pairs, a pathway analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang *et al.*, 2007). DAVID is a publically available high-throughput functional annotation tool and a gene list was created using the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (Table 3.1).

From this early analysis, the T1D-MVPS were associated with genes in which several are known to be related with immunity and T1D. For example, the HLA class I gene HLA-DQA2 (Husain *et al.*, 2008), HLA class II HLA-DOA (Santin *et al.*, 2009) and HLA-DPB2 (Lie *et al.*, 1999) have been linked to susceptibility to T1D. In the CD19 B cell dataset, several genes were listed that are involved in the insulin signalling pathway such as KRAS, GRB2 and PRKAR1B. Several MVPs for each cell type are listed in the appendix (Appendix III, Table 8.2, 8.3 and 8.4). Although there are genes listed that are associated with T1D, more twin pairs need to be ascertained to obtain more significant results.

Table 3.1. Functional annotation chart for $CD4^+$, $CD14^+CD16^-$ and $CD19^+$ cells.

Term	Count	P value	Genes
$CD4^+$			
Focal adhesion	23	0.001341	TNXB, FLT1, COL4A1, TLN2, DIAPH1, FLT4, BCAR1, ITGB4, ITGA11, PIP5K1C, ITGA3, ACTN3, MAPK10, SRC, MYL9, PAK6, LAMA2, CCND1, TNR, MAPK3, COL11A2, SHC2, MYLK
Hypertrophic cardiomyopathy (HCM)	13	0.002109	SLC8A1, PRKAG2, ITGA11, ITGB4, ITGA3, TTN, CACNA2D3, CACNA2D2, TGFB2, LAMA2, SGCD, RYR2, CACNA1C
Arrhythmogenic right ventricular cardiomyopathy	12	0.002621	LAMA2, SLC8A1, ITGB4, ITGA11, RYR2, GJA1, SGCD, ITGA3, ACTN3, CACNA2D3, CACNA1C, CACNA2D2
Regulation of actin cytoskeleton	23	0.003172	FGFR2, INS-IGF2, DIAPH1, MAP2K2, MRAS, ITGAE, BCAR1, ITGB4, ITGA11, PIP5K1C, ITGB2, ITGA3, ACTN3, MYL9, PAK6, KRAS, ITGAX, CHRM3, CHRM2, MAPK3, CYFIP1, CSK, MYLK
Axon guidance	16	0.004364	ABLIM2, NGEF, NTN4, LRRC4C, EPHB3, EPHB4, EPHB1, SLIT3, PAK6, SEMA6A, KRAS, ROBO1, MAPK3, SEMA4C, SEMA4B, NFATC1
Neurotrophin signalling pathway	15	0.007529	IRAK2, MAP2K2, FOXO3, MAPK10, YWHAE, TP73, YWHAG, KRAS, RPS6KA1, MAPK3, NTRK2, SH2B2, CSK, SHC2, NGF
Dilated cardiomyopathy	12	0.011311	LAMA2, SLC8A1, ITGB4, ITGA11, RYR2, SGCD, ITGA3, CACNA2D3, TTN, CACNA1C, CACNA2D2, TGFB2
Neuroactive ligand-receptor interaction	24	0.012421	OPRM1, GPR156, C3AR1, GABRG3, THRB, GABRA6, ADCYAP1R1, PTH1R, GABBR1, OXTR, FPR3, GABBR2, ADORA1, VIPR2, SCTR, GH1, PRLR, CHRM3, GRM8, GRIA1, CHRM2, P2RY2, TUBB3, GRID1
Prostate cancer	11	0.022978	E2F1, FGFR2, CCND1, HSP90AA1, KRAS, MAP2K2, INS-IGF2, MAPK3, TGFA, CREB5, RB1
Non-small cell lung cancer	8	0.026866	E2F1, CCND1, KRAS, MAP2K2, MAPK3, TGFA, RB1, FOXO3
Circadian rhythm	4	0.031014	CRY2, CSNK1E, PER1, BHLHE41
ECM-receptor interaction	10	0.039425	LAMA2, SDC1, TNXB, COL4A1, TNR, ITGB4, HSPG2, ITGA11, ITGA3, COL11A2
NOD-like receptor signalling pathway	8	0.051644	CCL11, NOD2, HSP90AA1, MAPK3, MAPK10, TRIP6, NLRP3, CCL7
Chronic myeloid leukemia	9	0.051707	E2F1, CCND1, BCR, KRAS, MAP2K2, MAPK3, RB1, SHC2, TGFB2
Glioma	8	0.055501	E2F1, CCND1, KRAS, MAP2K2, MAPK3, TGFA, RB1, SHC2
Cell adhesion molecules (CAMs)	13	0.057189	PTPRC, NRXN2, NRXN3, NFASC, ITGB2, NRXN1, HLA-DQA2, PDCD1, CDH5, SIGLEC1, SDC1, CNTNAP2, HLA-DOA
Bladder cancer	6	0.07778	E2F1, CCND1, KRAS, MAP2K2, MAPK3, RB1
Viral myocarditis	8	0.09271	LAMA2, CCND1, MYH15, SGCD, ITGB2, HLA-DOA, MYH8, HLA-DQA2
Pancreatic cancer	8	0.098153	E2F1, CCND1, KRAS, MAPK3, TGFA, MAPK10, RB1, TGFB2

$CD14^{+}CD16^{-}$			
Chemokine signalling pathway	5	0.025638	CXCL1, PRKCZ, ADCY2, JAK3, PXN
$CD19^{+}$			
Endocytosis	15	0.004792	PRKCZ, CLTA, RAB5B, RAB5C, MET, HLA-A, HLA-C, PIP5K1A, IGF1R, SMAP2, SH3GLB1, RAB11B, STAM, PARD6G, EHD2
MAPK signalling pathway	19	0.005298	FGF8, GRB2, RELB, NF1, TAOK3, DUSP10, PTPRR, MKNK1, CACNG2, MAP3K7, RPS6KA5, BDNF, ATF4, KRAS, RASGRP2, PPP3CB, PPP3CC, CACNA1C, IL1A
Glioma	7	0.02203	IGF1R, KRAS, GRB2, CAMK2D, CDK6, RB1, SHC4
Spliceosome	10	0.031686	PPIL1, SNRPB2, SNRPD1, SNW1, ACIN1, HNRNPC, SFRS1, DDX5, DDX42, BAT1
Melanoma	7	0.036998	IGF1R, FGF8, KRAS, MET, MITF, CDK6, RB1
Natural killer cell mediated cytotoxicity	10	0.042608	MICB, KRAS, ULBP3, GRB2, HLA-A, NFAT5, PPP3CB, PPP3CC, HLA-C, SHC4
Insulin signalling pathway	10	0.046144	PRKCZ, PRKAR2A, EIF4E, KRAS, GRB2, PRKAR1B, MKNK1, SH2B2, SOCS4, SHC4
Homologous recombination	4	0.072847	RAD51L1, MUS81, RAD51L3, RAD52
Axon guidance	9	0.081551	ABLIM1, EPHA6, KRAS, ROBO1, CFL2, MET, NFAT5, PPP3CB, PPP3CC
Long-term potentiation	6	0.088461	ATF4, KRAS, CAMK2D, PPP3CB, PPP3CC, CACNA1C
Prostate cancer	7	0.090522	HSP90AB1, IGF1R, ATF4, KRAS, GRB2, NKX3-1, RB1

3.3.7 At least 50 twin pairs are needed for a statistical significant result

The power of a study is important, particularly in human studies as with a small sample size, the study lacks power (Fitzner and Heckinger, 2010; Jones *et al.*, 2003). As there were no overlapping differences between the cell types and initial findings showed small differences between the twin pairs, the study was underpowered. The top 100 significant differences found in $CD4^+$ T cells were used to produce a power curve (Figure 3.7). A power calculation was performed under the assumption that some of the differences found in the experiment were real (true positives) but are being swamped by other non-true differences (false positives). To increase the power of the study, the significance of the true positives must increase and the significance of the false positives must decrease. To do this, the number of twins must increase or the technical noise must be reduced. The technical noise was determined to be very low, therefore the number of samples have to be increased.

It was calculated for 50 twin pairs to give 90% power for the top DMP, 64% power for the top 50 DMPs, 56% power for the top 100 DMPs. To obtain 90% power for the top 50 and 100 DMPs, at least 70 MZ pairs discordant for T1D would be needed. From this analysis, it was decided to asertain more twins to a total of 50 pairs discordant for T1D.

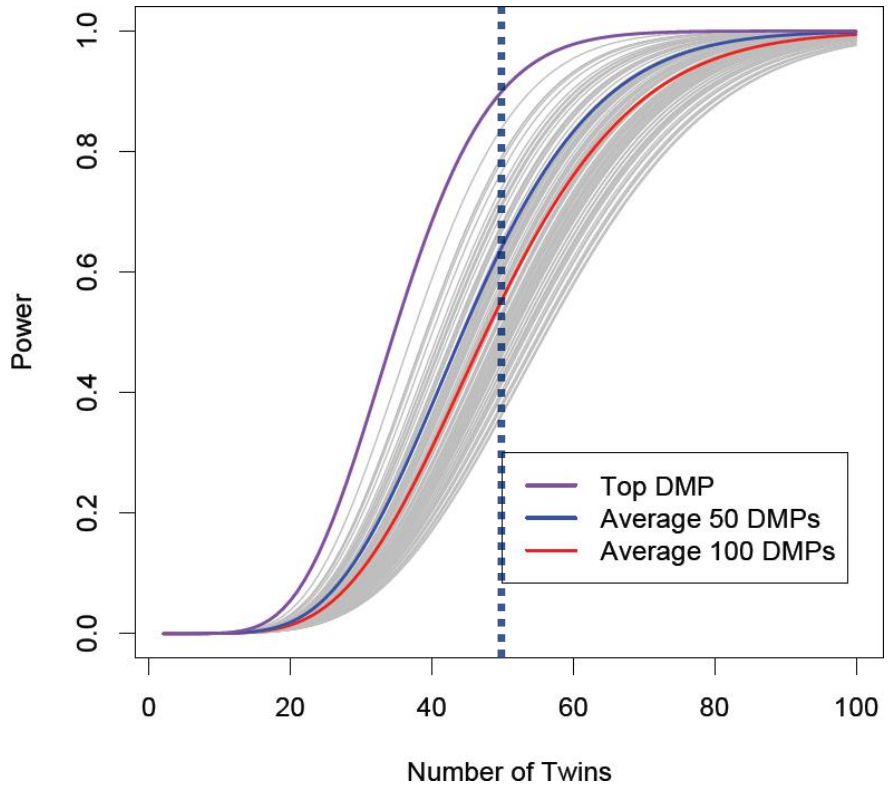


Figure 3.7. Power curves against number of twins. The top DMP (purple curve) has the most power at 90% for 50 twin pairs, and also the lowest number compared to the top 50 DMPs (blue curve) and top 100 DMPs (red curve). At 50 twin pairs, the top 50 DMPs would give 56% power and for the top 100 DMPs, 64% power, which was low.

3.3.8 More hypermethylated regions in BS-seq data compared to Illumina450K

To gain a more comprehensive look at methylation in the genome, further analysis was performed with BS-seq. To compare the datasets from Illumina450K and BS-seq, $CD4^+$ samples from four pairs of twins that were also profiled in the Illumina450K array were analysed. Analysis was performed using *bsseq* in R and resulted in an average of 3.5 reads per CpG (median ~5). The overall mapping efficiency was high at approximately 80% for the samples and duplication for these libraries was also quite low, on average less than 1%. Methylation levels for each CpG site were measured using beta values. The beta value is the ratio of the methylated probe intensity and the overall intensity and range from completely unmethylated (0 or 0%) to completely methylated (1 or 100%) on a continuous scale (Du *et al.*, 2010).

CpG sites that were covered by more than 10 reads across all samples were compared to the 450K sites in the same individual, twin and unrelated individual (Figure 3.8). The regions from the BS-seq data were significantly different to that of the Illumina450K data. In the same individual, there were more hypermethylated CpG sites than the Illumina450K (Figure 3.8A). The same was seen when comparing a twin pair (Figure 3.8B) and the one individual with an unrelated individual (Figure 3.8C). As expected, when comparing a twin pair with just Illumina450K data, there was a high correlation (Figure 3.8D). The high number of hypermethylated sites from the BS-seq data, may have been due to coverage issue. Therefore only CpG sites that had 30x coverage were analysed (Figure 3.9). By doing so, the majority of the hypermethylated sites were removed however there was still a trend of hypermethylation on BS-seq compared to Illumina450K. Whilst the correlation was high when comparing a twin pair using the Illumina450K dataset, when comparing the same sites in the twins analysed by BS-seq and unrelated individuals, there was still significant noise over these sites (Figure 3.10A and B). Further investigation revealed that there was no correlation between the differences called between the Diabetic vs Control from the 450K calls against the Diabetic vs Control from the BS-Seq data (Figure 3.10C).

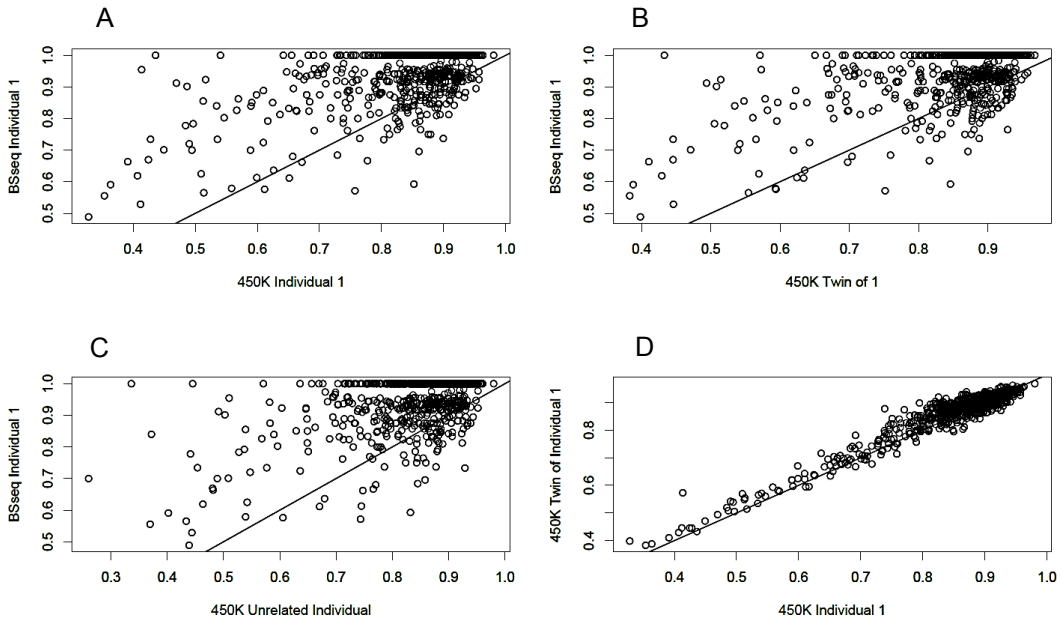


Figure 3.8. Comparison of BS-seq and Illumina450K datasets. Correlation plot of CpG sites from all samples from the BS-seq compared to the 450K sites in the same individual (A), Twin 1 against individual 1 (B) and unrelated individual and Individual 1 (C). 450K calls were compared with Individual 1 against the twin of individual 1 (D).

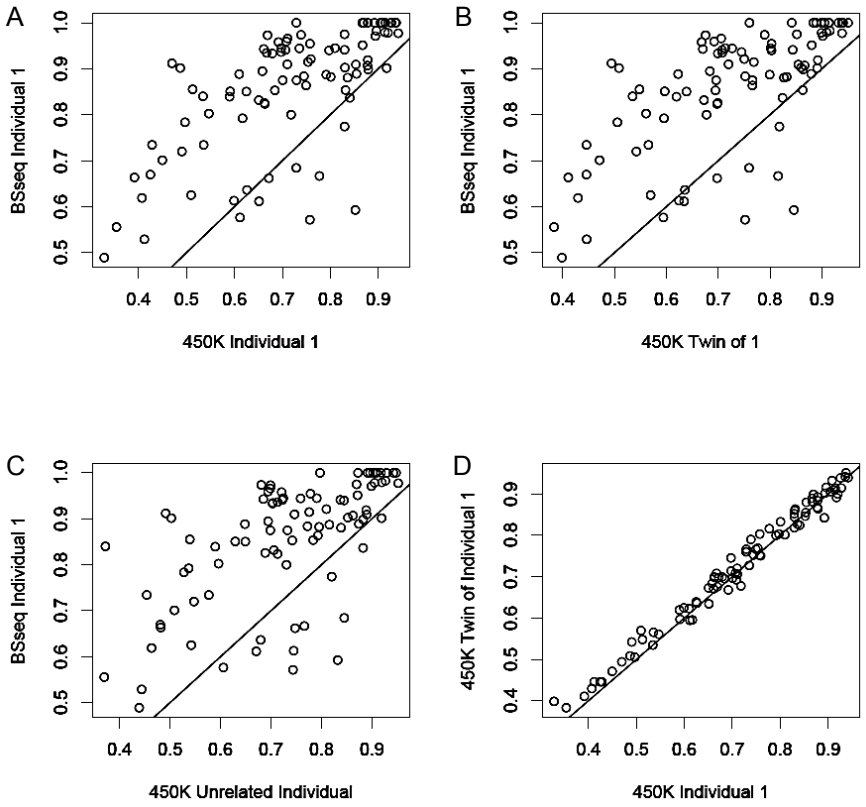


Figure 3.9. Clean up of BS-seq data at 30x coverage. Correlation plot of CpG sites from all samples from the BS-seq compared to the 450K sites in the same individual (A), Twin 1 against individual 1 (B) and unrelated individual and Individual 1 (C). 450K calls were compared with Individual 1 against the twin of individual 1 (D). There was less noise, however there was still a trend of hypermethylation on BS-seq.

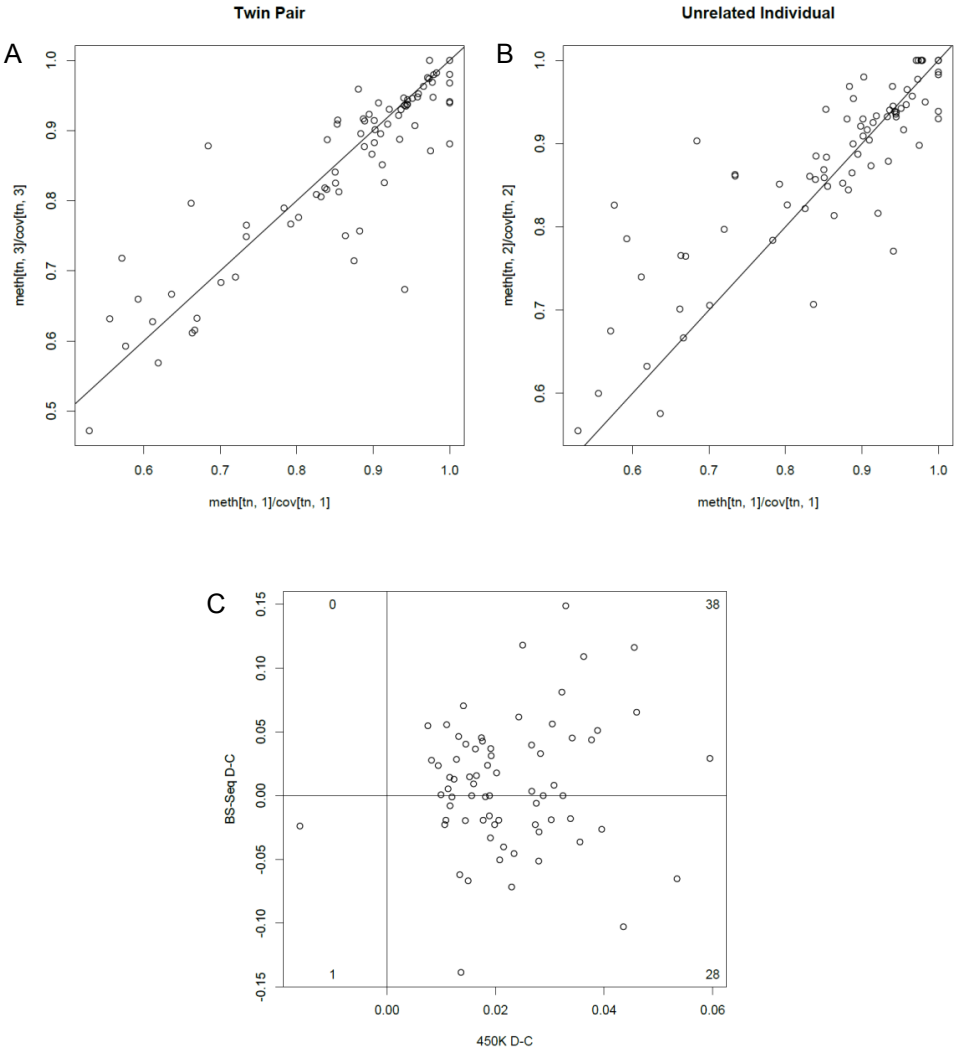


Figure 3.10. Further comparisons of BS-seq data at 30x coverage in CD4$^+$ cells. Comparison of the same sites from 450K data on BS-seq data from a twin pair (A) and unrelated indivudal (B). There was still significant noise over these sites. Differences between diabetic and non-diabetic calls from Illumina450K compared to difference between diabetic and non-diabetic calls from BS-seq were also plotted (C). There was no significant correlation.

3.3.9 DMR sets overlapped $CD4^+$ specific enhancer sites

Once it was established that better coverage was required in order to find significant differences, the eight $CD4^+$ samples were sequenced a second time. This time to try and identify new DMPs that existed outside of the regions not covered by Illumina450K, as there are approximately 27 million CpG sites in the human genome (McClay *et al.*, 2014) and Illumina450K only assesses approximately 485,000 (Sandoval *et al.*, 2011), whereas BS-seq analyses all of the CpG sites (Li and Tollefsbol, 2011).

Sequencing the samples the second time yielded minimum coverage of 5x compared to approximately 3.5x the first time the samples were sequenced. Data points with low coverage were omitted from the analysis leaving approximately 20 million CpGs at median coverage approximately 14x. There was a possible twin effect seen at the CGI positions (Figure 3.11A) and CpG sites (Figure 3.11B) investigated. An example of a DMR was shown in Figure 3.11C. Further differential methylation analysis yielded 125 DMRs (Table 3.1). Most of which were found in CD4 enchancer sites marked by H3K4me1.

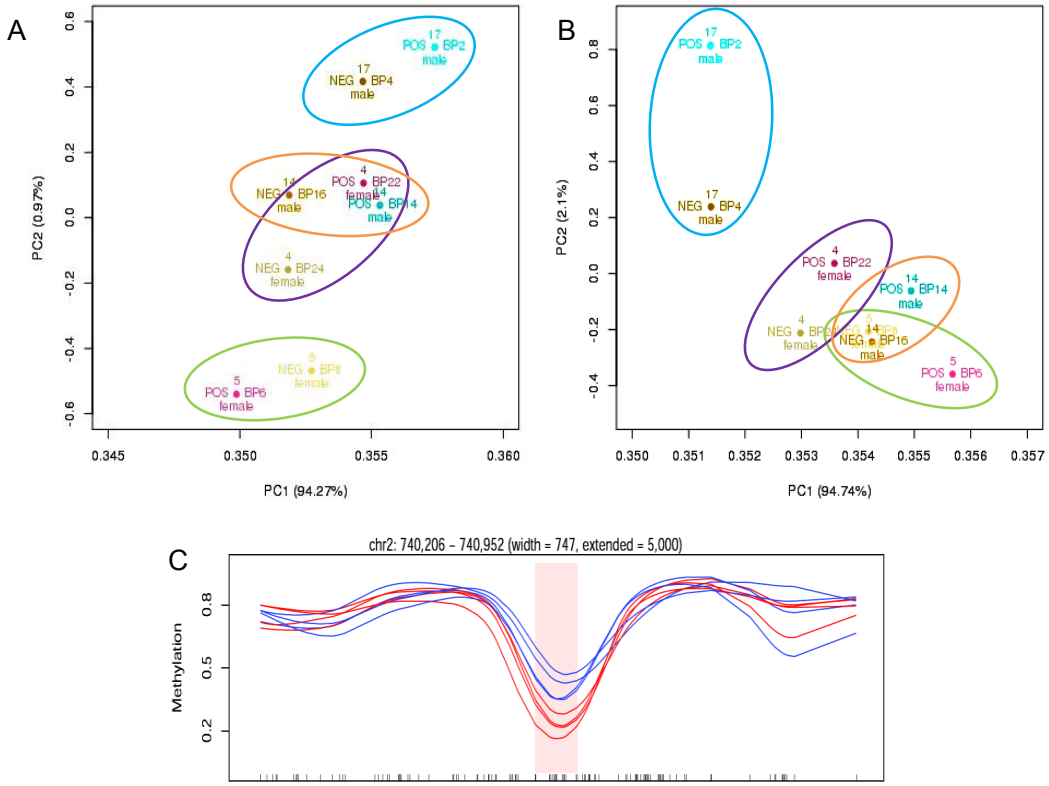


Figure 3.11. Coverage at CpG islands and CpG sites. PCA over T1D data at CpG islands (A) and CpG sites (B). There could be a possible twin effect. An example of a DMR with the diabetic samples in red and the control in blue (C). The DMR is highlighted in red. At this particular DMR, hypomethylation was observed in the diabetic twin compared with the healthy co-twin.

Table 3.2. Differentially methylated regions across different genomic features.

Feature	Total	Hyper	Hypo	% Hyper	% Hypo	% Total
CGI	17	12	5	70.59	29.41	13.60
DHS	44	26	16	59.09	36.36	35.20
Exons	33	17	16	51.52	48.48	26.40
Intergenic	37	22	15	59.46	40.54	29.60
5'UTR	3	2	1	66.67	33.33	2.40
Known_genes	79	39	40	49.37	50.63	63.20
LADS	8	8	0	100.00	0.00	6.40
Lines	17	5	12	29.41	70.59	13.60
lncRNA	6	5	1	83.33	16.67	4.80
Promoters	57	28	29	49.12	50.88	45.60
Repeats	49	24	25	48.98	51.02	39.20
Shores	60	29	31	48.33	51.67	48.00
CD4_enhancers	115	64	51	55.65	44.35	92.00
TFBS	90	53	37	58.89	41.11	72.00
3'UTR	8	4	4	50.00	50.00	6.40
Introns	69	33	36	47.83	52.17	55.20
Genome	125	69	56	55.20	44.80	100.00

3.4 Discussion

3.4.1 Significant correlation between preliminary study T1D-MVPs and discovery cohort

This study investigated DNA methylation differences between MZ twin pairs discordant for T1D. MZ twin pairs are particularly useful in epigenetic studies as genetic variation is a confounding factor which was eliminated in this study. The similarities in methylation differences seen between the twin pairs from the preliminary cohort and the same pairs in the present cohort can serve as a validation study as it has same cohort profiled using different technologies. The fact that the Illumina450K array was able to detect significant differences between the diabetic and non-diabetic twin pairs in the same cell type, $CD14^+$ monocytes, suggests that the DNA methylation signatures are stable.

Rakyan *et al*. also focussed on promoter-associated single CpGs (Rakyan *et al.*, 2011a). This limitation was due to the use of the Illumina27K which interrogated CpG sites mapped to the promoter regions of genes (Bibikova *et al.*, 2009). This present study was able to focus on genomic regions outside of the promoter region such as CGI shores and other regions flanking them, providing a more comprehensive methylome coverage.

3.4.2 Cord blood in determining causal effects of methylation

In epigenetic studies, it is difficult to determine whether a significant DNA methylation difference is a cause or consequence of a disease. It has been reported that T1D-MVPs appear before diagnosis of T1D in autoantibody positive individuals (Rakyan *et al.*, 2011a), therefore a useful cohort to study would be individuals without autoantibodies but then go on to develop T1D. If the T1D-MVPs were not present in these individuals, then these MVPs may arise as a result with the appearance of autoantibodies, and in the case of childhood onset diabetes, early seroconversion to autoantibodies can occur from nine months to two years of age (Ziegler *et al.*, 2012; Parikka *et al.*, 2012). If the T1D-MVPs were enriched in the individuals with no autoantibodies present, this indicates that the T1D-MVPs antedate the immune process which leads to T1D. Analysing methylation marks in cord blood provides an insight to whether these epigenetic marks were present in utero and therefore antedate disease. DNA methylation marks have also been

shown to be stable for at least three years from birth (Beyan *et al.*, 2012a). The group extracted genomic DNA from Guthrie cards to asses of in utero-derived DNA methylation variation. Inter-individual DNA methylation variation was identified at birth and three years later, suggesting that disease-relevant epigenetic variation could be detected at birth.

However, with small differences identified in the sorted cells in this present study, it would be hard to detect in unsorted samples such as cord blood. However, the cell type composition in cord blood is different compare to adult whole blood (Beck and Lam-Po-Tang, 1994). For example, the percentages of cell types reported in the paper were (cord blood vs adult blood): CD3 (57.7% vs 74.1%, CD16 (18.7 vs 6.6%) and αβTCR (56.3 vs 71.1%). Although there was a weak correlation between the cord blood samples and the T1D-MVPs (from $CD14^+$ monocytes) identified in the preliminary study, follow up samples from the same individuals from whom the cord blood was obtained from, will be collected. This is useful as later analyses comparing cord blood with T1D-MVPs identified in the present study will be more informative as there were more cell types studied interrogating over 450,000 CpG sites.

3.4.3 Differentially methylated positions are cell type-specific

A novel finding from this study was that the methylation changes seemed to be cell type-specific. DNA methylation has already been shown to be tissue-specific (Rakyan *et al.*, 2008) and Rakyan *et al*. had studied one cell type in a later study and identified 132 different CpG sites that were differentially methylated in MZ twins (Rakyan *et al.*, 2011a). Here, four cell types were profiled using the successor of the Illumina27K. Over 485,000 CpGs were analysed compared to just the 27,000. From this analysis, a total of 258 DMPs were identified in the Illumina450K array, 170 of which were detected in $CD4^+$ T cells, followed by the $CD19^+$ B cells with 79, $CD14^+CD16^-$ with 8 and buccal cells with 1 DMP. It was assumed that there would be some overlap between the four cell types. This was the case in one study investigating DNA methylation in SLE. A research group had assessed the role of DNA methylation in $CD4^+$ T cells, $CD19^+$ B cells and $CD14^+$ monocytes also using the Illumina450K platform (Absher *et al.*, 2013). They had collected blood samples from 49 SLE patients and 58 controls and identified methylation differences in 166 CpGs in $CD19^+$ B cells, 97 in $CD14^+$ monocytes and 1,033 in $CD4^+$ T cells. The authors also reported differences that were common to all the cell types which are in

contrast to the findings of this study, where the DNA methylation differences were cell type-specific.

Following on from this, a preliminary functional analysis was performed. Several genes identified from the KEGG pathway are known to be associated with T1D. This included HLA class I HLA-DQA2 (Husain *et al.*, 2008), HLA class II HLA-DOA (Santin *et al.*, 2009) and HLA-DPB2 (Lie *et al.*, 1999) and KRAS (Gout *et al.*, 2013). Rakyan *et al*. had reported hypermethylation in TNF and hypomethylation in GAD2 and HLA-DQB1 in twin pairs discordant for T1D (Rakyan *et al.*, 2011a) and Stefan *et al*. reported hypomethylation in HLA-E and hypermethylation in CD226 and HLA-DQA2 (Stefan *et al.*, 2013), which was also identified in this study.

3.4.4 Importance of individual cell populations in EWAS

There are DNA methylation studies investigating the methylome using PBMCs (Li *et al.*, 2010). PBMCs are easy to extract and can provide useful information on the epigenomic landscape. However, analysis in whole blood or even PBMCs could be problematic as each cell type has its own methylation profile (Reinius *et al.*, 2012; Adalsteinsson *et al.*, 2012). Therefore it was necessary to isolate and purify specific immune cells for this study and clear clustering of the different cell types was sown (Figure 3.1C), apart from the buccal samples which suggest there were a mixture of cell types that could have contaminated the samples.

Correcting for cell type composition bias is important (Lowe and Rakyan, 2014). Through the use of bioinformatics, it is possible to distinguish different leukocyte subsets in whole blood (Accomando *et al.*, 2014; Koestler *et al.*, 2013). Although employing a bioinformatics-based approach in identifying the different leukocyte subsets in whole blood, there are subsets within each leukocyte. For example, T cells have subset including T_H cells and T_{regs} (Figure 3.12). Identifying individual cell methylome would be interesting as epigenetic regulation plays a role in normal immune cell differentiation. Also, the balance between processing time and cost per sample for sorting cells must be considered in a study.

As previous analyses were based on reference datasets, Houseman et al have developed a method for conducting EWAS analysis when a reference dataset is unavailable (Houseman *et al.*, 2014). This is particularly useful when reference datasets from tissues such as placenta, saliva or tumour tissue are unknown.

Different groups are actively developing algorithms in order to predict and assess proportions of leukocytes in whole blood samples. If significant DNA methylation signatures in T1D emerge from this study, these T1D-MVPs will be looked for in whole blood, as this tissue is most readily available from any cohort. If the signature can be detected in whole blood, this reduces the cost and time of sorting particular cell types.

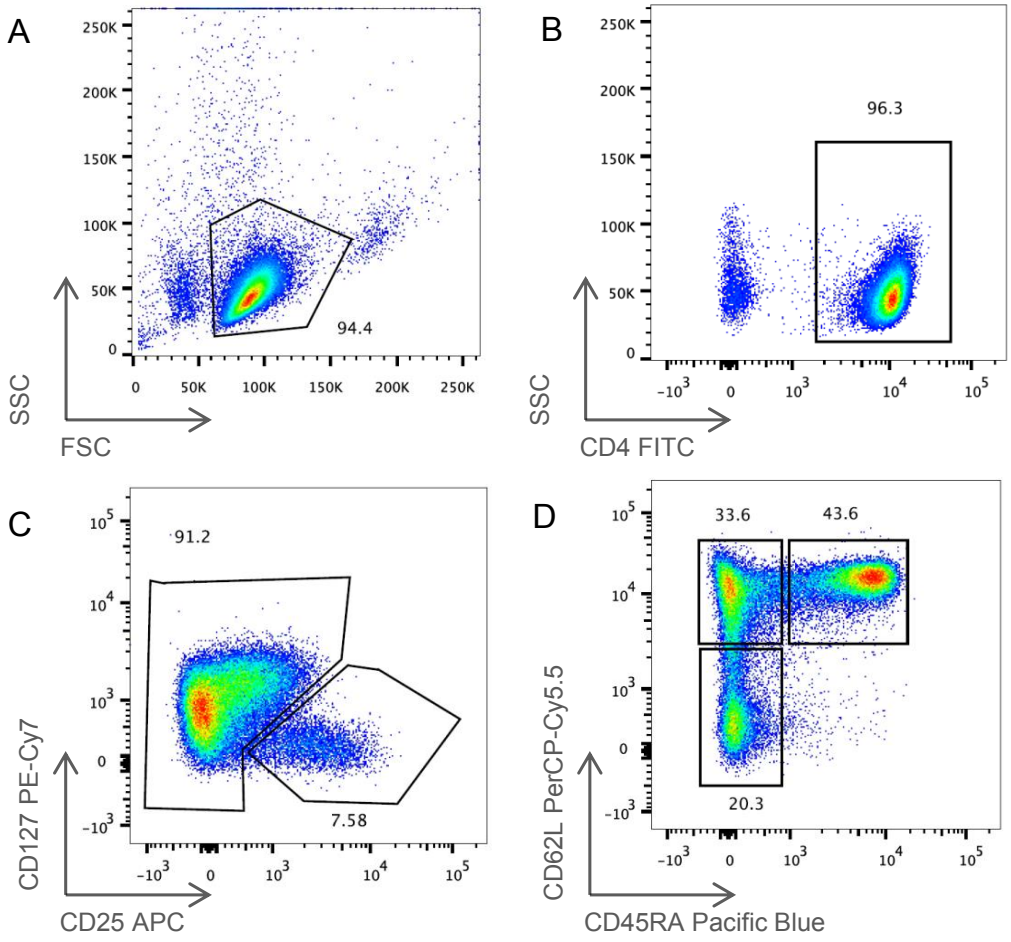


Figure 3.12. CD4$^+$ T cell subsets analysed by flow cytometry. CD4$^+$ T cells have several subsets. Within the lymphocyte population (A), CD4$^+$ cells (B) are further separated into memory (left gate) or T_{regs} (right gate) (C). Memory cells are then sorted into central memory T cells (top left), effector memory T cells (bottom left) and naïve T cells (top right) (D).

3.4.5 DMRs identified overlapped $CD4^+$ enhancers

Initially, BS-seq data on the CD4 T cells did not provide enough coverage for analysis. The samples were sequenced a second time providing more coverage. Differential methylation analysis generated 125 DMRs which were found in CD4 enhancers marked by H3K4me1 (ENCODE data). Enhancers are regions of DNA that enhances transcription levels of a gene through the binding of transcription factors (Hardison and Taylor, 2012). There are more than a million enhancers, therefore, many more than there are genes. Because of this, a number of genes are regulated by the same enhancer, which may co-localize with CpGs (Ziller *et al.*, 2013). As enhancers are important in cell type-specific regulation of gene expression (Sanyal *et al.*, 2012; de Laat and Duboule, 2013), differential methylation at these enhancer sites could play a role in the development of disease. Particular in $CD4^+$ T cells as T1D is believed to be a T cell mediated autoimmune disease.

Enhancers are modulated by a combination of transcription factors, DNA methylation and histone modifications (Blattler *et al.*, 2014). But do they correlate with alterations in other epigenetic marks such as histone modifications associated with T1D. There is evidence to suggest that histone modifications under diabetic conditions that could account for the progression or aetiology of the disease (Miao *et al.*, 2014; Miao *et al.*, 2012). Genetic factors are also important as GWAS SNPs were reported to be enriched in the regulatory DNA of disease-specific cell types (Maurano *et al.*, 2012). Therefore the integration of GWAS and EWAS data to study gene expression and the local chromatin architecture will help define how DNA methylation can impact molecular outcomes in different immune cells and therefore disease.

3.5 Conclusion

In conclusion, this study identified small but significant DNA methylation differences between 24 MZ twin pairs discordant for T1D. These epigenetic marks were found to be stable as they appeared in the same twin pairs sampled years before. This study used the Illumina450K array, the successor to Illumina27K, in which more CpG sites were interrogated, specifically outside of the promoter region. Different cell types have separate DNA methylation profiles therefore it was important to separate these cells from PBMCs, otherwise false positives results may be detected. The strongest methylation signal was detected in the $CD4^+$ T cells. Out of 258 DMPs, 170 were from the T cells, followed by the $CD19^+$ B cells with 79, $CD14^+CD16^-$ with 8 and buccal cells with 1 DMP.

Although, T1D-associated genes were identified in a preliminary functional analysis, more twin pairs are needed to obtain sufficient power in the study. DMRs were also found to overlap $CD4^+$ enhancer sites from the BS-seq analysis. Alterations to the chromatin by epigenetic modifications have been implicated in human disease which in turn can be used to identify potential therapeutic targets.

Chapter 4

Detection of unmethylated DNA in the human insulin gene for monitoring β-cell death

4 Detection of unmethylated DNA in the human insulin gene for monitoring β-cell death

4.1 Introduction

Insulin controls blood glucose levels and is expressed mainly in the β-cells (Melloul *et al.*, 2002), in which the baseline β-cell population is established before five years of age (Gregg *et al.*, 2012). In T1D patients, most of the β-cells are destroyed, however patients with disease duration for over 50 years, still show some signs of insulin production (Keenan *et al.*, 2010). In mice, it was shown that β-cell neogenesis occurs during embryogenesis and decreases shortly after birth, but neogenesis in murine adult pancreas does not occur (Xiao *et al.*, 2013). However, combined treatment of DA-1229 and $Pam3CSK_4$ could reverse T1D by increasing the number of replicating β-cells and β-cell mass (Kim *et al.*, 2012). In humans, Lebastchi *et al.* reported preservation of β-cell function in patients who received teplizumab, an anti-CD3 monoclonal .antibody, compared to the placebo (Lebastchi *et al.*, 2013).

Currently, C-peptide secretion is one of the available clinical biomarkers to measure residual β-cell function in T1D (Barker *et al.*, 2014). C-peptide produced is released in a 1:1 ratio with insulin after the cleavage of proinsulin (Wahren *et al.*, 2012). C-peptide measurement can be measured from serum samples to assess endogenous insulin reserve in diabetic individuals (Davis *et al.*, 2014). It has also been proposed that protein phosphatase 1, regulatory (inhibitor) subunit 1A (PPP1R1A) can be used as a novel real-time biomarker for acute β-cell destruction (Jiang *et al.*, 2013). Using immunoprecipitation, PPP1R1A was found to be released from injured islets and was proportionate to the extent of β-cell death. Free-circulating DNA have been studied for diagnostic purposes (Ziegler *et al.*, 2002) for cancer, stroke (Tong and Lo, 2006) and sepsis (Rhodes *et al.*, 2006) as higher levels of circulating DNA have been found in patients compared to the matched controls. In T1D, it has been reported that there were CpG sites that were uniquely unmethylated in the insulin gene promoter region, both in mice and humans (Kuroda *et al.*, 2009). The same group analysed the same CpG sites in eight other tissues including liver, lung and spleen, and found these sites to be mainly methylated (Husseiny *et al.*, 2014). Methylation of these unmethylated CpG sites in the insulin promoter suppressed gene activity by almost 90% (Kuroda *et al.*, 2009). Detection of unmethylated insulin DNA in serum has already been performed by different research groups who

reported differences in β-cell death between diabetic patients and normal controls (Husseiny *et al.*, 2014; Lebastchi *et al.*, 2013).

In this chapter, a method was developed to detect methylation differences between MZ twin pairs discordant for T1D in circulating DNA. The aim was to assess whether the insulin promoter CpG demethylation may play a crucial role in β-cell death. At first, tissue-specific unmethylation in the gene body (exon 2) of the inulin gene was investigated. The next step was to study DNA methylation in the promoter region of the insulin gene using methylation-specific PCR, previously described by other groups (Akirav *et al.*, 2011; Husseiny *et al.*, 2012). As mentioned previously, hypomethylated gene regions were found in insulin only in β-cells, therefore any β-cell destruction would release unmethylated DNA into the circulation, hence the use of a sensitive assay such as qPCR to detect DNA from serum. The final part investigated the use of pyrosequencing to detect methylation levels in the twin samples.

4.2 Aims and objectives

The aim of this study was to develop an assay to detect unmethylated CpG sites in the insulin gene. Primers were designed for qPCR to amplify certain unmethylated CpG sites, which have been reported to be tissue-specific, in the insulin promoter region. This assay was used to try to identify whether there was a difference in methylation between MZ twins in serum samples. DNA was extracted from the serum samples and then bisulfite converted to analyse methylation. Pyrosequencing was used later to sequence four CpG sites in the insulin promoter.

4.3 Results

4.3.1 Non-specificity of methylation state at exon 2 in the insulin gene

To assess methylation in insulin, a nested PCR assay was developed. Initially exon 2 of the insulin gene was studied as it was suggested that tissue-specific unmethylated CpG sites were located there (Lebastchi *et al.*, 2013; Akirav *et al.*, 2011). DNA was extracted from 300 µL of serum, bisulfite converted and used in a nested PCR assay. The first step of the nested PCR was performed in a thermal cycler. The PCR product was ran on an agarose gel, cleaned up and then used as a template for the second step, which was performed in the ABI Real-Time 7500 system. The primer sets in this reaction discriminated between unmethylated and methylated CpG sites +273 and +399 in exon 2. The qPCR products were confirmed by agarose gels and DNA sequencing. Samples from healthy controls were used for the dilution series (Figure 4.1A and D) and melting curves were performed (Figure 4.1B and E). Primer efficiencies for the methylated and unmethylated primers were 92% and 91% respectively (Figure 4.1C and F).

Once the amplification efficiency testing was performed, no difference was seen between the two primer sets (Figure 4.2A). Different primer concentrations were used to try and optimise the assay. The different concentrations were 0.25 µM, 0.5 µM, 1 µM, 1.5 µM and 2 µM. Nevertheless, there was still no difference between the two primer sets (Figure 4.2B). Samples from a twin pair discordant for T1D and a healthy control were used to try to differentiate the two primer sets as it has been reported that methylation levels differ between T1D patients and healthy controls (Lebastchi *et al.*, 2013; Husseiny *et al.*, 2014). All samples showed non-specificity between the methylated and unmethylated DNA (Figure 4.2C).

In summary, non-specificity between methylated and unmethylated sites was a problem in this assay hence this method could not be used to monitor β-cell death in T1D patients.

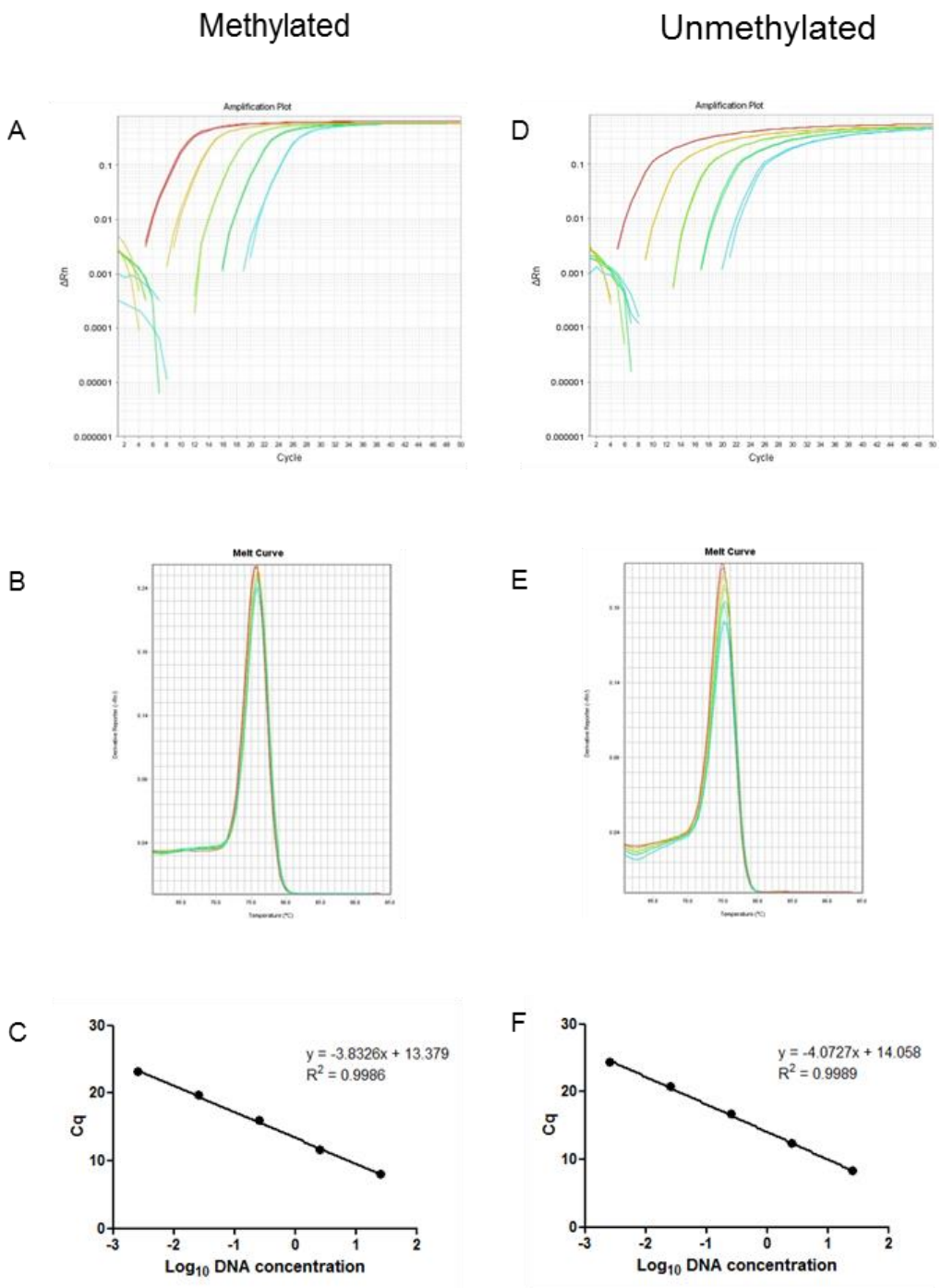


Figure 4.1. Amplification and melting curves for the methylated and unmethylated primer sets. Amplification curves of the primer against the methylated CpG sites (A), melting curve (B) and standard curve (C). Amplification curves of the primer against the unmethylated CpG sites (D), melting curve (E) and standard curve (F). The single peaks in the melting curve plots represent a single specific product amplified. Primer efficiencies should be within 10% of each other for a high efficient assay.

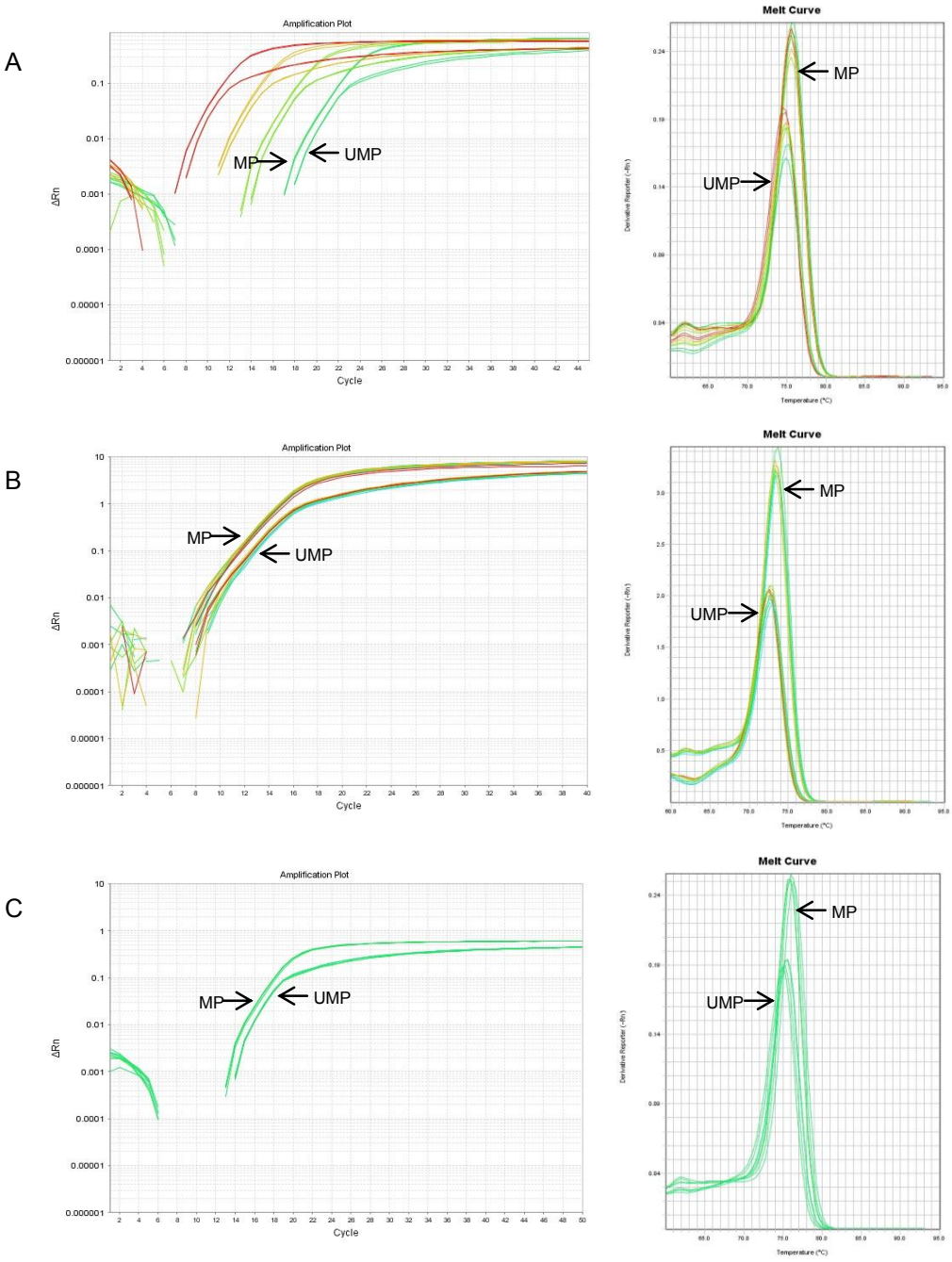


Figure 4.2. Optimisation of qPCR revealed non-specificity between methylated and unmethylated CpG sites. Different optimisation steps showed non-specificity issues with the primer sets. There was no difference in specificity between the two primer sets with the 10x dilution series (A). There was no difference in specificity using different primer concentrations (B) and there was no difference when comparing samples from a twin pair discordant for T1D and a healthy control (C). All melting curve plots showed two specific products that were amplified from the methylated and unmethylated primer sets.

4.3.2 Unmethylation at CpG sites in the insulin promoter region

As non-specificity between the methylated and unmethylated CpG sites was an issue in exon 2, the attention turned to the promoter region of the insulin gene. There are nine CpG sites in the promoter region compared to exon 2 (8), intron 1 (2) and intron 2 (2) (Husseiny *et al.*, 2014). In this section, four of the nine CpG sites were studied. In this assay, a nested PCR was performed also, however both steps were carried out in the ABI Real-Time 7500 system. Three sets of primers were designed as described previously (Husseiny *et al.*, 2014), this time interrogating four CpG sites at -234, -206, 135 and -36 relative to the transcription start site (TSS). These four CpG sites were previously described to be unmethylated compared to other tissues such as lung, liver and spleen (Husseiny *et al.*, 2014).

A plasmid with the insulin gene insert was grown and purified for the serial dilutions to determine amplification efficiency and to use as a positive control in this assay. An amplification efficiency test was performed and showed that the BSP primers produced primer dimers (Figure 4.3). The appearance of the primer dimers were confirmed by gel electrophoresis (Figure 4.4). Primer dimers are an issue as SYBR Green binds to any double-stranded DNA. This made it difficult to determine whether the signal was coming from the primer dimer or the unmethylated CpG sites. Several optimisation steps were carried out for the BSP primer set only, as MSP1 and MSP2 did not need optimising. The concentrations of the primers were optimised by performing a primer optimisation matrix (Table 4.1). The appearance of primer dimers when using the BSP primers, resulted in no clear distinction of the signal between the samples and NTC (Figure 4.5).

Different SYBR Green master mixes were also used as the master mixes contains different concentrations of magnesium chloride (Figure 4.6). Magnesium chloride assists the reaction as a cofactor for DNA polymerase, therefore the more magnesium chloride there is, the more powerful the reaction. The three master mixes used were Power SYBR Green (ABI), MESA Blue (Eurogentec) and QuantiTect SYBR Green (Qiagen). Primer dimers still appeared with all master mixes. Other optimisation steps included using DMSO (Appendix III, Figure 8.3) and shorter primers (Appendix III, Figure 8.4), but the primer dimers still occurred.

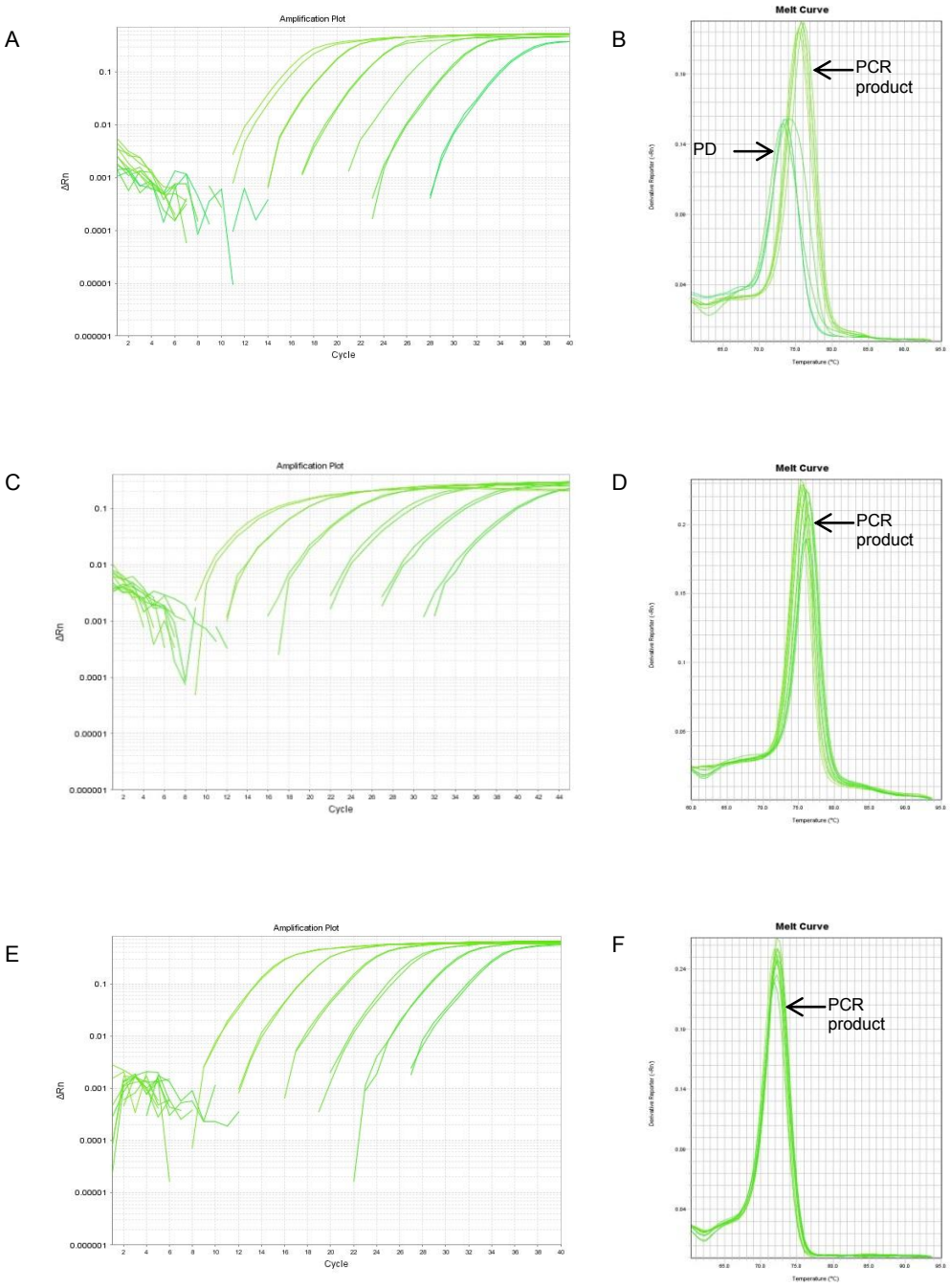


Figure 4.3. Amplification and melting curves for the three primer sets. Amplification and melting curves for the BSP (A and B), MSP1 (C and D) and MSP2 (E and F) primer sets are shown. The melting curve of the BSP primer set shows another peak other than the PCR product, this was a primer dimer effect. Primer dimers have a lower dissociating temperature as the products tend to be smaller than the amplicon of interest. The melting curves of the MSP1 and MSP2 primers set show one distinct peak, therefore there were no primer dimers. PD: primer dimer.

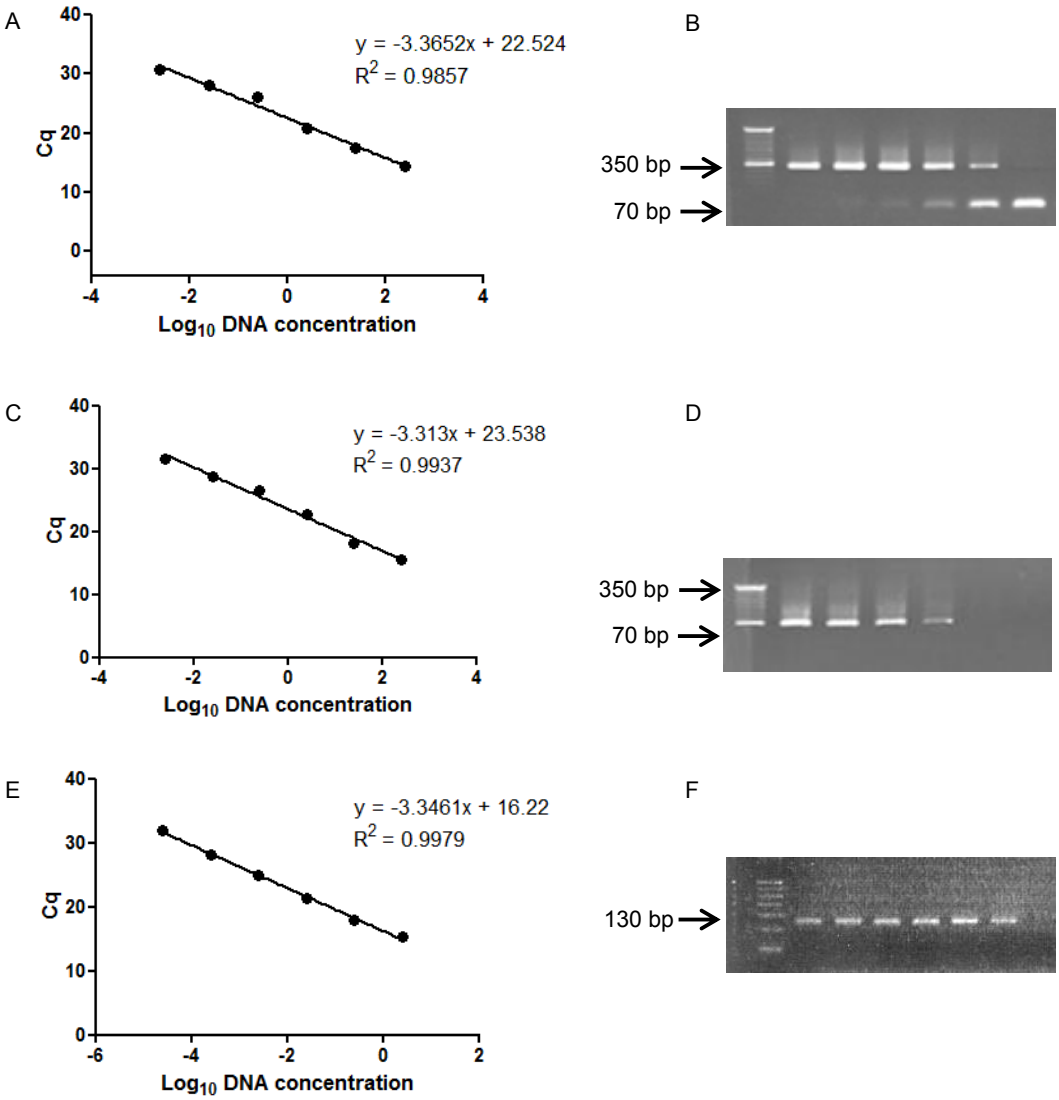


Figure 4.4. Primer efficiency and specificity of the three primer sets. Standard curves and PCR products visualised on an agarose gel for BSP (A and B, MSP1 (C and D) and MSP2 (E and F). Primer efficiency for each primer set were BSP 98.5±3.4%, MSP1 98.7±4.1% and MSP2 90.5±6.6%. Primer dimers were seen with the BSP primer set at approximately 70 bp. PCR products sizes were BSP: ~350 bp, MSP1: ~350 bp and MSP2: ~130 bp.

Table 4.1. Primer optimisation matrix.

Forward primer (nM)	Reverse primer (nM)		
	100	200	400
100	100/100	100/200	100/400
200	200/100	200/200	200/400
400	400/100	400/200	400/400

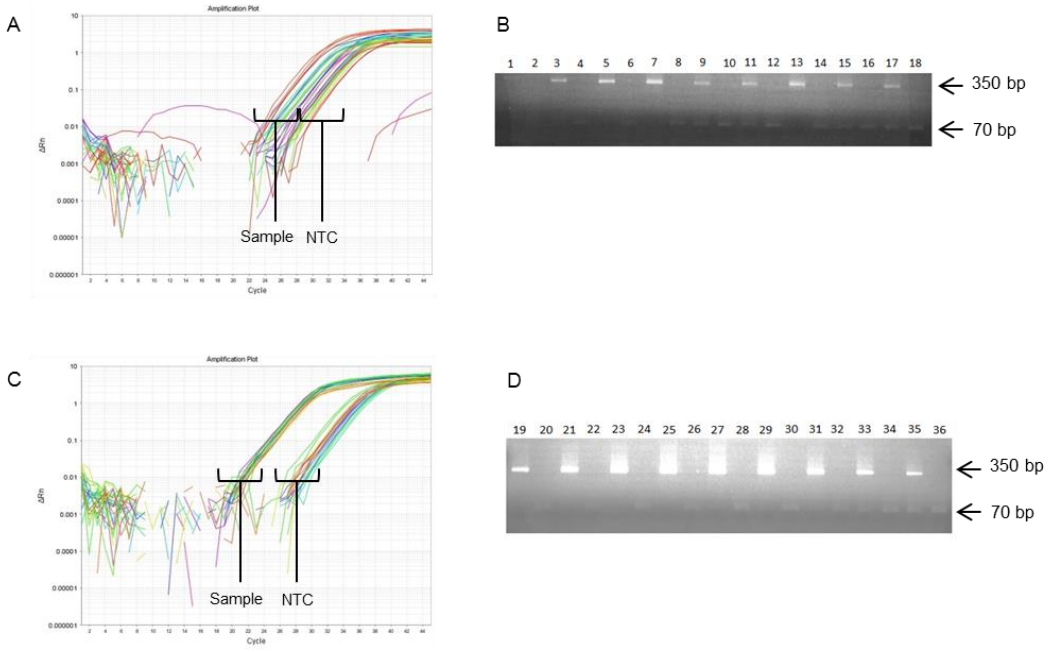


Figure 4.5. Optimisation steps of methylation specific PCR. There was no distinct signal between the samples and NTC (A) when running with BSP primers. The primer dimers also showed up after running the products on a agarose gel (B). the primer dimer length came up to approximately 70 bp. Whereas when running with MSP1 primers, there was a distinct difference between the samples and NTC (C) and no primer dimers were seen until high concentrations of the primers were used (D).

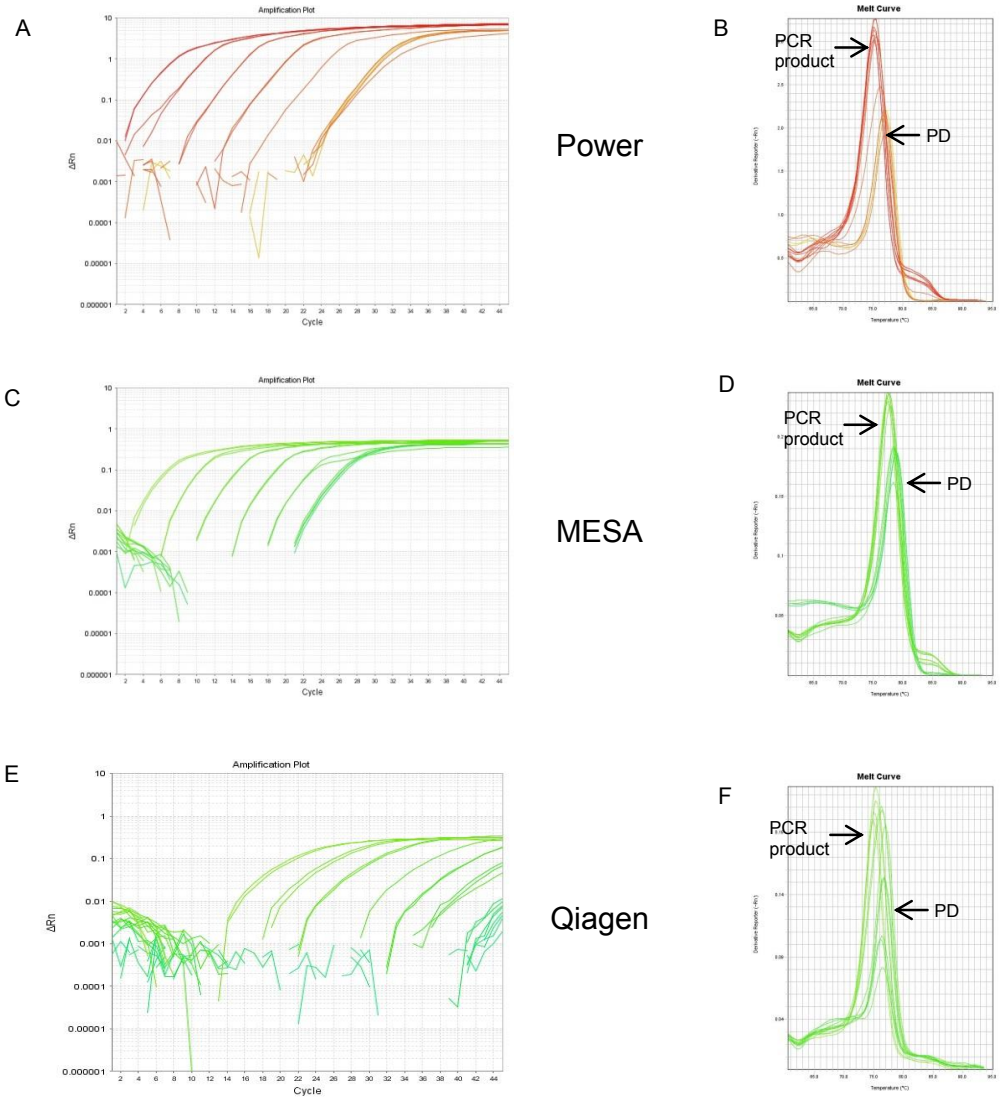


Figure 4.6. qPCR reaction with different SYBR Green master mixes. Different SYBR Green master mixes were used to eliminate primer dimers from the assay. Power SYBR Green from ABI (A and B), MESA Blue from Eurogentec (C and D) and QuantiTect SYBR Green from Qiagen (E and F). Primer dimers occurred still with the different master mixes. PD: primer dimer.

4.3.3 Primers could distinguish between methylated and unmethylated templates

To assess whether the primer sets could discriminate between methylated and unmethylated CpG sites, fully methylated and unmethylated plasmids with the insulin promoter was used in a qPCR reaction (Figure 4.7). The plasmids were fully methylated with M.SssI CpG methyltransferase, which methylated all cytosines residues.

As expected, the BSP primers did not discriminate between methylated and unmethylated plasmids with similar Cq values (Figure 4.7A). The Cq differences between the unmethylated and methylated plasmid were 1.780±0.6893. The melting curve showed two distinct peaks (Figure 4.7B), which was expected as two different products were being amplified. The MSP1 primer set was able to distinguish the methylated and unmethylated plasmid, showing lower Cq for the unmethylated plasmid than the methylated plasmid (Figure 4.7C). This was also expected as the primer set was designed to amplify only unmethylated CpG sites. The Cq differences between the unmethylated and methylated plasmid were 4.768±0.5581. The melting curve showed two distinct peaks (Figure 4.7D). The MSP2 primer set was also able to distinguish the methylated and unmethylated plasmid, showing a lower Cq for the unmethylated plasmid than for the methylated plasmid (Figure 4.7E). The Cq differences between the unmethylated and methylated plasmid were 7.150±1.740. The melting curve showed two distinct peaks within one another (Figure 4.7F). This may be that both products were much smaller than the previous two primer sets. BSP and MSP1 both amplified a product or approximately 350 bp whereas MSP2 amplified a product approximately 130 bp.

In summary, the BSP primer set did not discriminate between the methylated and unmethylated CpG sites however, the MSP1 and MSP2 did, as expected. Although the primers could detect unmethylated CpG sites of interested, the primer dimers made it difficult to determine whether the signal is coming from the CpG sites or the primer dimers therefore this assay cannot be used as a quantitative assay to monitor β-cell death.

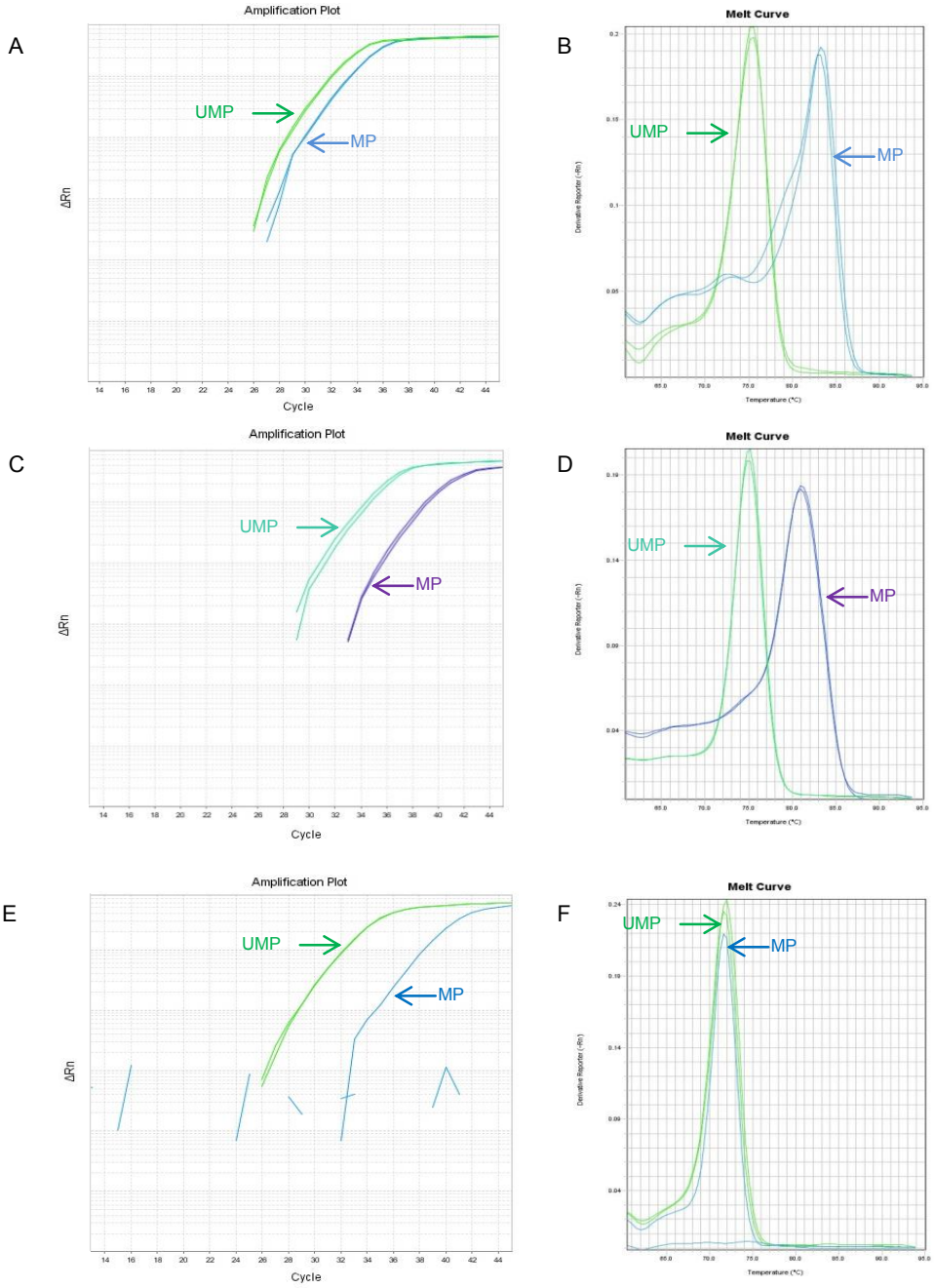


Figure 4.7. Methylation analysis in fully methylated and unmethylated plasmids. There was no discrimination between the methylated and unmethylated plasmid with the BSP primer set (A). There were two distinct peaks in the melting curve from the two different products the unmethylated and methylated plasmids produced (B). MSP1 primer set was able to distinguish between the methylated and unmethylated plasmid (C and D), and similar results was shown from the MSP2 primer set (E and F).

4.3.4 Taqman assay was highly specific

The Taqman assay was performed to eliminate primer dimers as this technology is highly specific. High specificity of the assay is due to the additional probe included in the reaction. The probe is an oligonucleotide that has a fluorescent reporter dye attached to the 5' end and a quencher on the 3' end. During each cycle, the probe is intact and the quencher reduces the fluorescence emitted by the reporter dye, therefore no fluorescence is detected by the machine. However, when a product is amplified, the probe breaks down, increasing the fluorescence of the reporter dye. The detection is more specific than non-probe based methods such as SYBR Green, which binds to any double-stranded DNA, therefore when the amplification efficiency assay was performed, the Cq values came up higher than those when using SYBR Green. Because of the high Cq values, a 5x dilution series was performed instead of a 10x dilution series (Figure 4.8A and B). A new primer set (BSP2), MSP2 and a probe was designed for this assay (Table 2.1). Primer efficiencies with Taqman were: BSP2 79% (Figure 4.8C) and MSP2 88% (Figure 4.8D). The PCR products were then run on an agarose gel (Figure 4.8Eand F).

As in Section 2.13.5.1, DNA was extracted from the serum samples and treated with sodium bisulfite. Unlike the nested PCR assay, Taqman was carried out as a one-step assay as the two-step assay did not yield any results (data not shown). The assay was performed with a healthy control sample, however, the signal for MSP2 was lower than the signal for BSP (Figure 4.9). This result was in contrast to what was expected as the BSP primer set did not discriminate between methylated and unmethylated CpG sites, hence more product should have been amplified compared to the MSP2 primer set. As the Taqman assay was highly specific and did not amplify as much DNA as SYBR Green, pyrosequencing was explored.

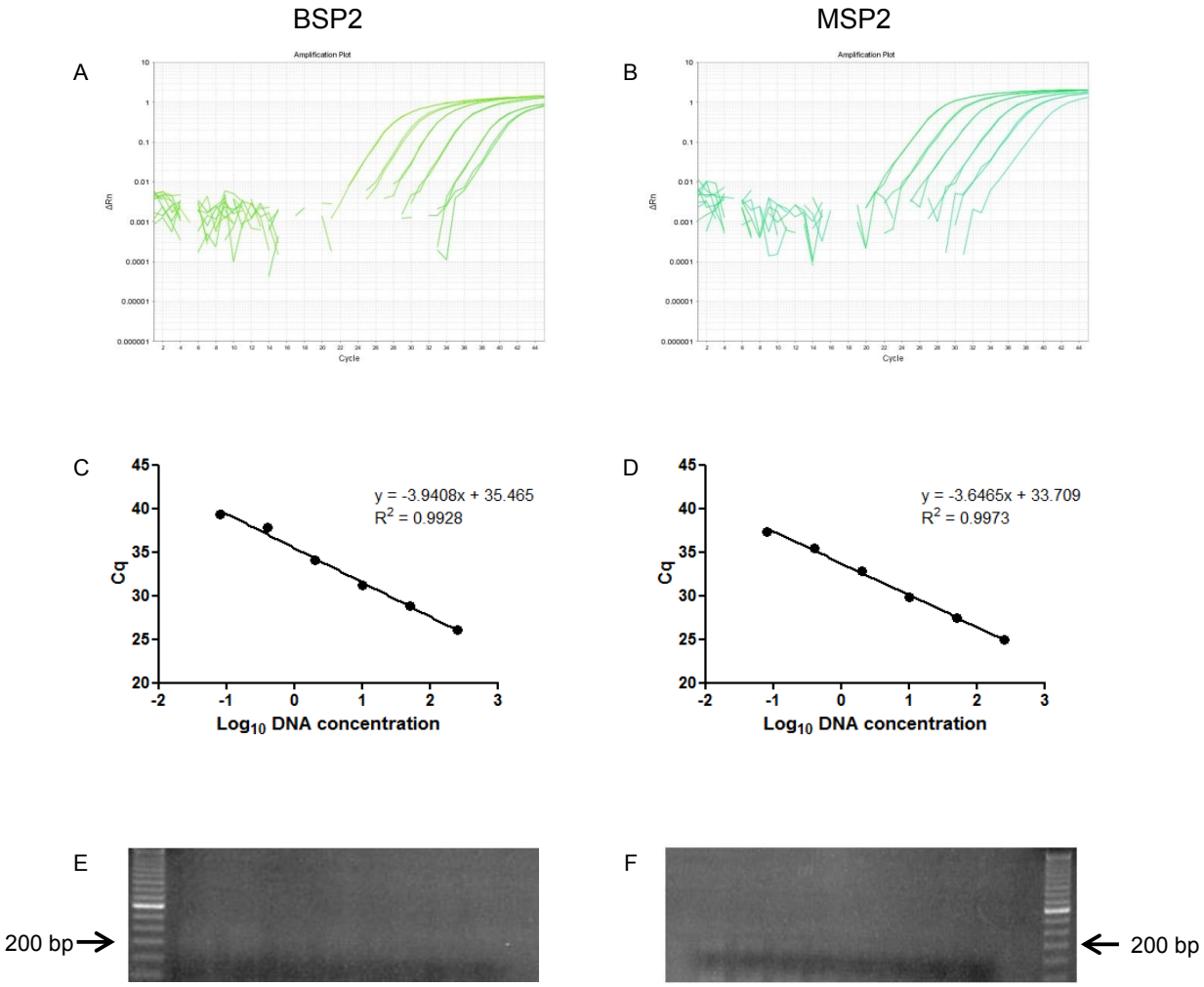


Figure 4.8. DNA methylation analysis using the Taqman technology. Amplification curves showing with BSP2 (A) and MSP2 (B) primer sets. Standard curves of BSP2 (C) and MSP2 (D) primer sets. Gel images of BSP2 (E) and MSP2 (F) PCR products. The PCR products were approximately 200 bp and no product was seen for the NTC samples.

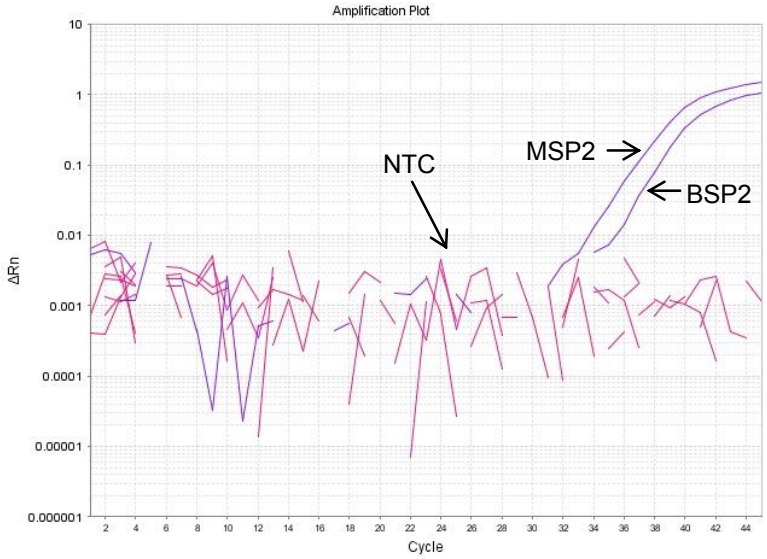


Figure 4.9. MSP2 signal was lower than BSP2 from a control sample. A serum sample from a healthy control was analysed using the Taqman technology. The signal for MSP2 was lower than BSP2 which was unexpected. There was no signal for the NTC samples.

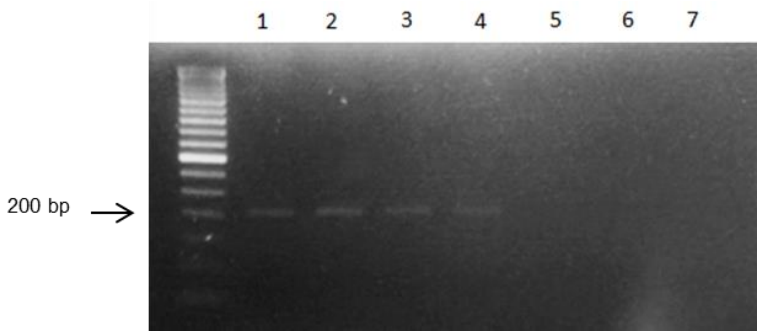
4.3.5 Pyrosequencing

4.3.6 Optimisation of the PCR step

Pyrosequencing is a sequencing-by-synthesis method, where the incorporation of a single nucleotide causes a chain of reactions that produces visible light. Each light that is produced, determines which nucleotide was incorporated into the DNA sequence. The assay was performed to compare the methylation levels between the MZ twin pairs using serum samples. First, sample preparation before pyrosequencing had to be optimised. A gradient PCR was performed to establish the optimum annealing temperature (Figure 4.10A). Once the optimum annealing temperature was determined, PCR was performed with forward and reverse primers (Figure 4.10B), with the reverse primer biotinylated (Figure 4.11). The PCR products were loaded into a 96-well plate and then sent off to the Genome Centre for sequencing.

This method was used to sequence short regions to interrogate the CpG sites which had been previously assessed by qPCR in Section 4.3.3 and 4.3.4. Two sequencing primers were designed, SP1 and SP2 (Figure 4.11). The sequencing primers also interrogated an additional two CpG sites which have been reported to also be unmethylated, at -180 and -102 (Husseiny *et al.*, 2014).

A



B

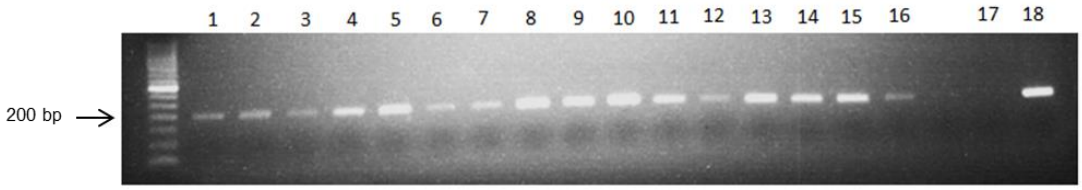


Figure 4.10. Optimisation of PCR step. Once DNA had been treated with sodium bisulfite, a gradient PCR was performed to establish the optimum annealing temperature (A). A 50 bp ladder was loaded, annealing temperatures from Lanes 1 to 7 were 53.6°C, 54.9°C, 56.5°C, 59.1°C, 62.2°C, 65.9°C and 70.3°C. The optimum annealing temperature was taken at 55°C. The optimised annealing temperature was then used in the cycling steps for PCR (B). A 50 bp ladder was loaded. The primers amplified the DNA (Lanes 1 to 16). A negative control was loaded into Lane 17 and a positive control was loaded into Lane 18. The positive control was the plasmid that was purified in Section 2.13.1.

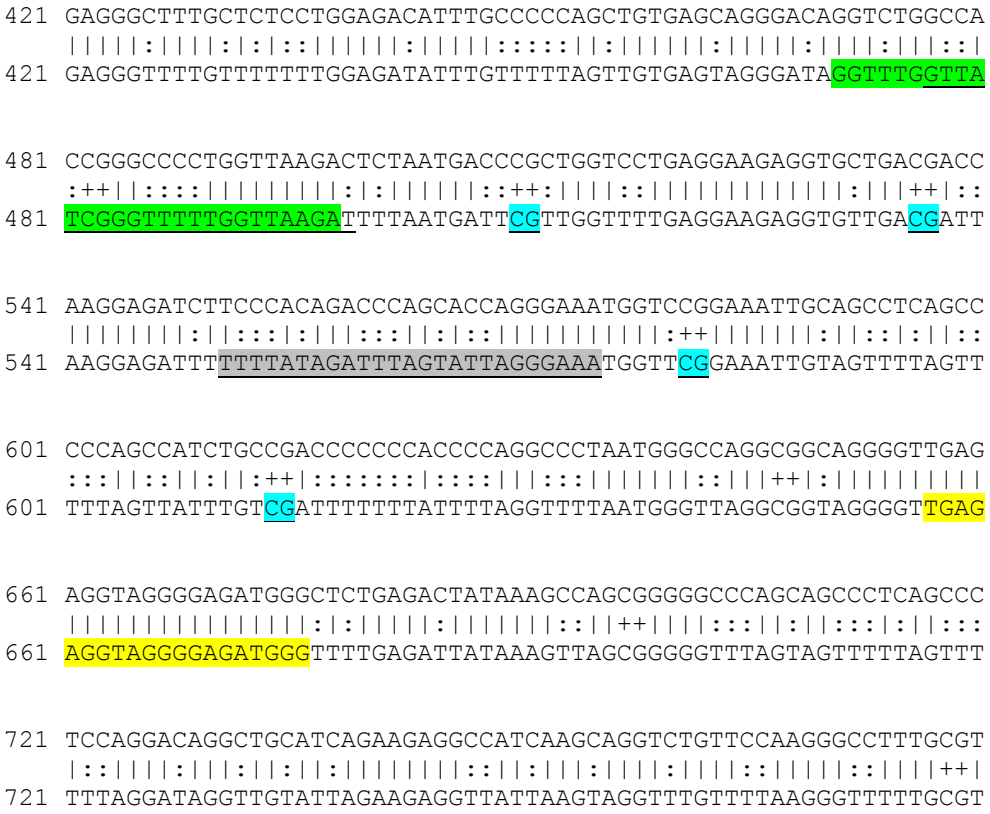


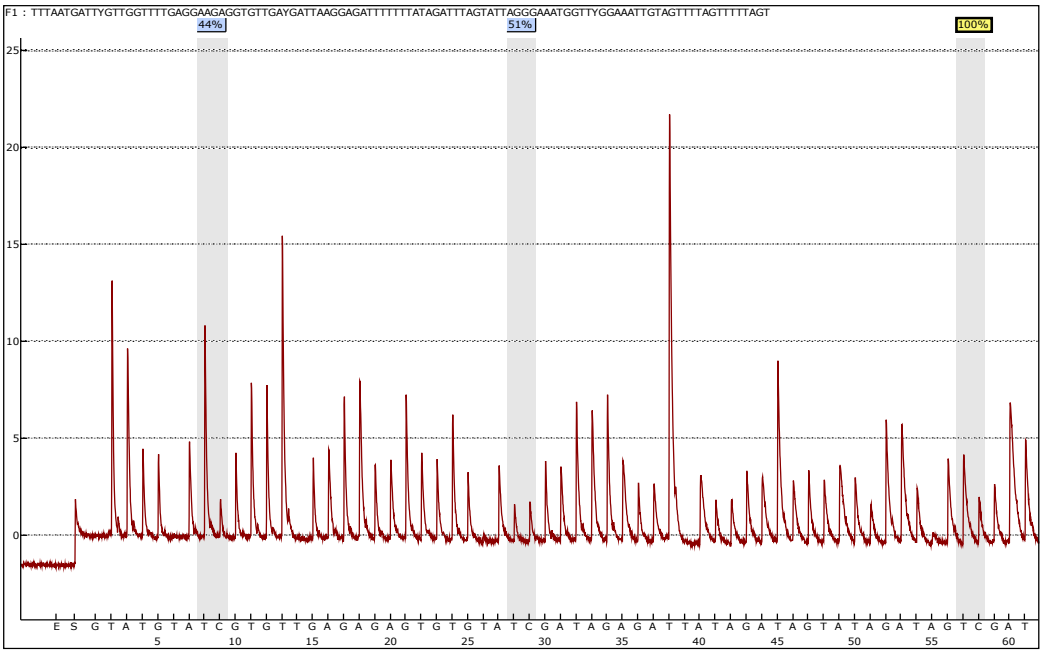
Figure 4.11. Primer design of PCR and sequencing. Primers were designed using PyroMark (Qiagen, UK). The forward primer (green) and reverse primer with biotin (yellow) were used to produce an amplicon approximately 200 bp. Sequencing primer SP1 (underlined) included CpG sites -206 and -180 and SP2 (grey and underlined) included CpG sites -135 and -102. CpG sites of interested are highlighted in blue and underlined.

4.3.7 Comparison of intra twin pair methylation

Once the PCR step was optimised, the PCR products were sent off to the Genome Centre for sequencing. Samples from 33 twin pairs discordant for T1D were analysed, producing pyrograms (Figure 4.12). Two sequencing runs were performed as the two CpG sites of interest were too far apart to sequence at one time. The first sequencing primer (SP1) interrogated CpG -206 and the second sequencing primer (SP2) interrogated CpG -135. From the pyrogram, SP1 sequenced three CpG sites, with the first, the CpG of interest (Figure 4.12A). The second was CpG site -180 and the third was CpG site -135. The optimum amplicon for pyrosequencing tends to be between 80 bp to 150 bp. Although SP1 had sequenced three CpG sites, the last CpG site was omitted from analysis as sample checks had failed towards the end of the amplicon. The same applied to SP2, as it also sequenced three CpG sites (Figure 4.11). Therefore, in addition to CpG sites -206 (CpG 1) and -135 (CpG 3), CpG sites -180 (CpG 2) and -102 (CpG 4) were also analysed due to the sites being sequenced. Each sample produced a methylation percentage (Appendix III, Table 8.5 and 8.6).

Several pairs of the 33 pairs analysed were omitted from the final analyses due to the samples failing checks during sequencing. This was due to the low amount of DNA present initially. Hence for SP1, a total of 24 pairs were analysed and for SP2, a total of 23 pairs were analysed. Individual methylation levels were plotted for each CpG site (Figure 4.13). Methylation levels were higher for the twin pairs at CpG sites 1 and 3 compared to CpG sites 2 and 4. Mean values for CpG 1 in T1D 80.23±16.15%, Non-T1D 83.61±12.15%, CpG 2 T1D 61.71±13.57%, non-T1D 65.17, CpG 3 73.45±13.79%, non-T1D 76.26±12.17% and CpG 4 T1D 59.19±11.93, non-T1D 64.40±13.55. Majority of the intra-pair methylation differences (non-diabetic – diabetic) were small (Figure 4.14).

A



B

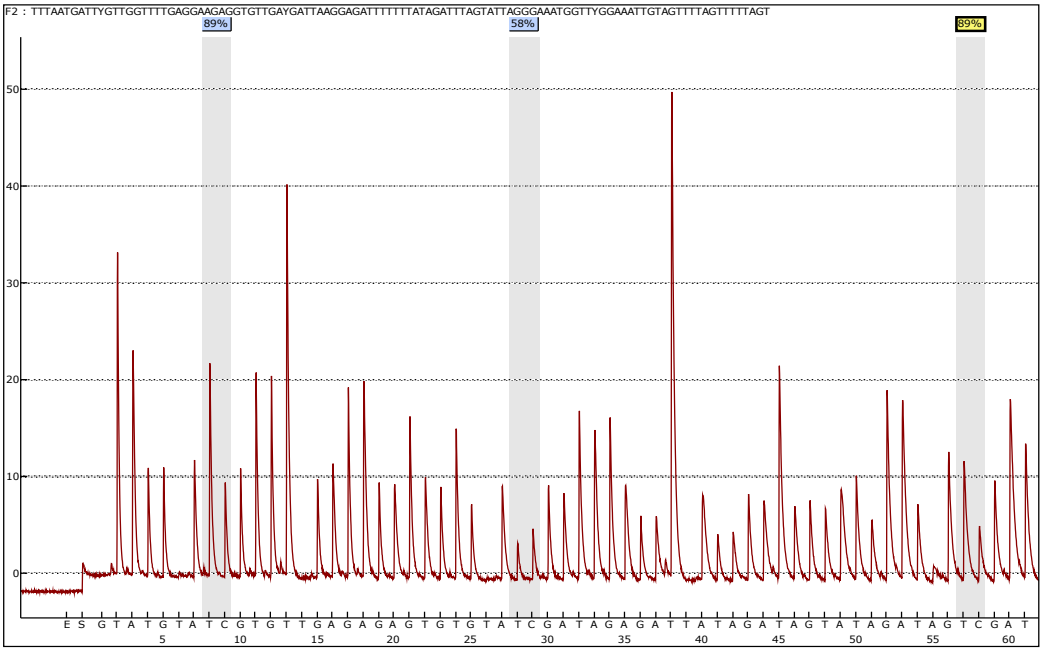


Figure 4.12. Pyrogram of CpG sites in the insulin promoter region in a twin pair. Pyrogram from the diabetic twin (A) and non-diabetic twin (B). The CpG sites are highlighted by the grey shaded area. Methylation values are boxed above the CpG sites.

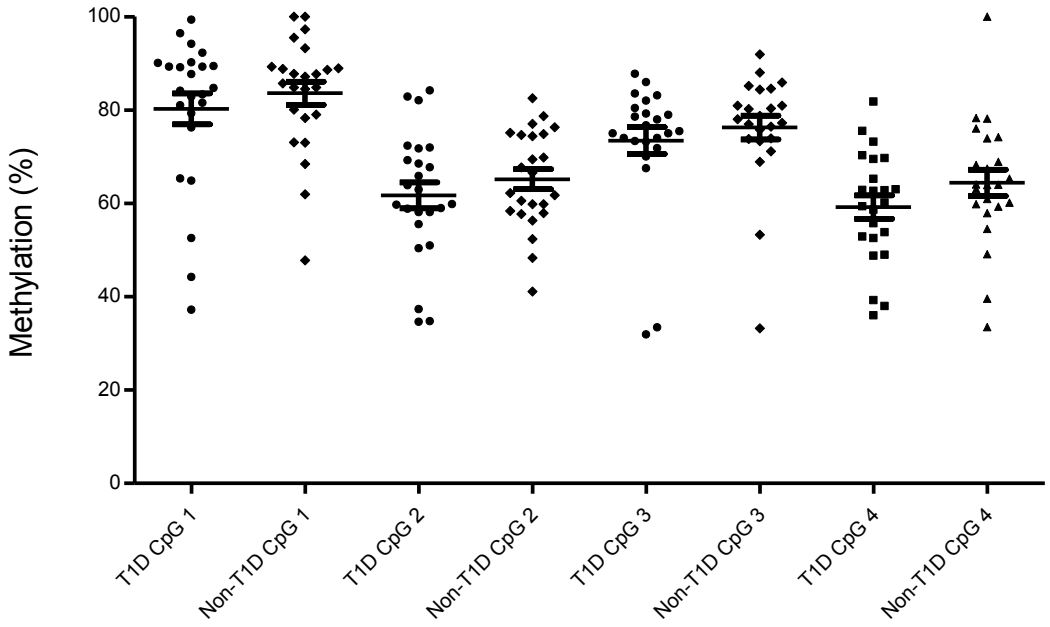


Figure 4.13. Methylation levels of each individual at the four different CpG sites. Methylation levels were plotted for 24 twin pairs at CpG 1 and 2, 23 pairs at CpG 3 and 4. No significant difference was seen between the T1D and non-diabetic co-twin at each CpG site.

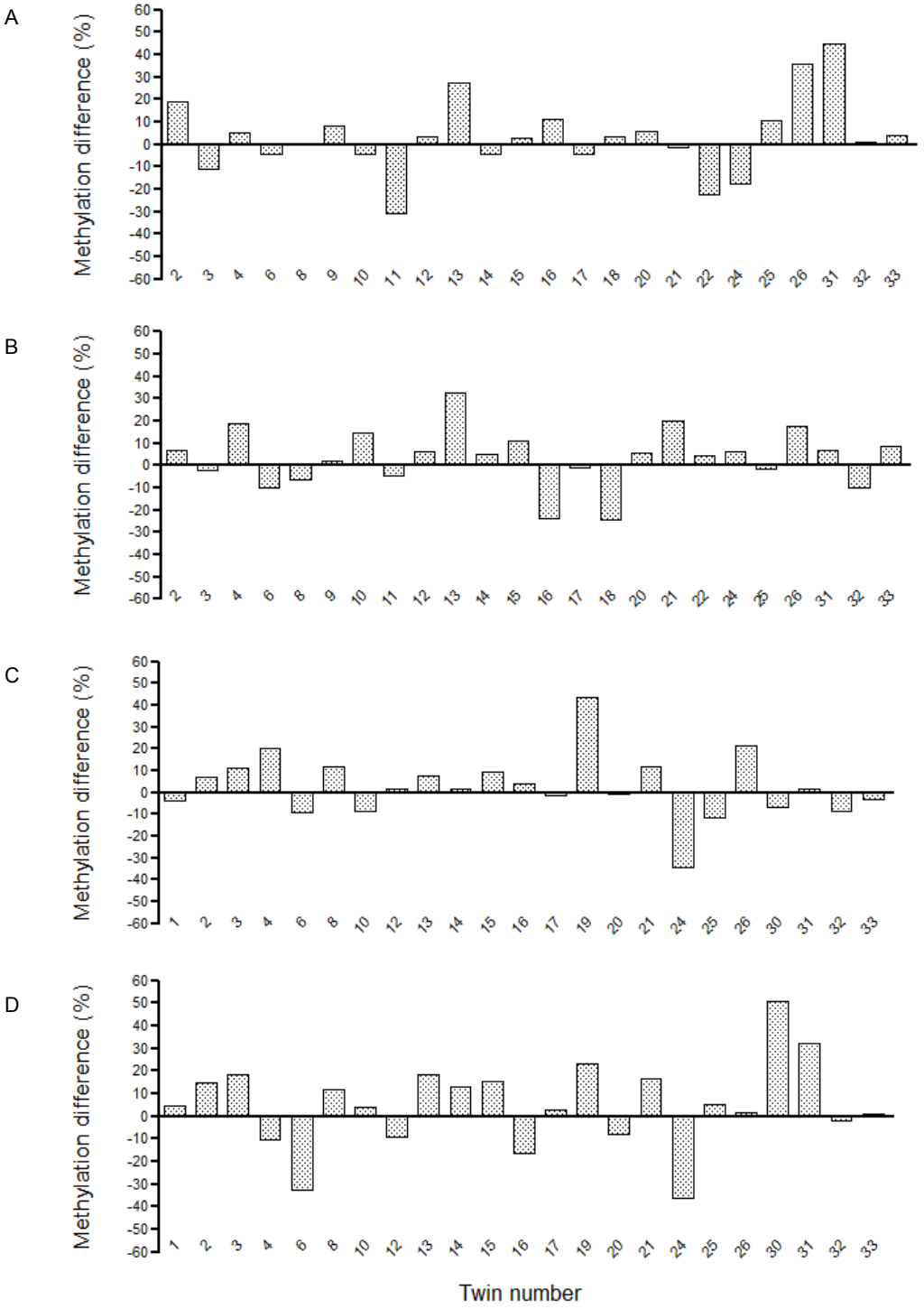


Figure 4.14. Methylation differences between MZ twin pairs. The methylation differences between each twin pair were plotted (Non-diabetic – diabetic) at CpG 1 (A), CpG 2 (B), CpG 3(C) and CpG 4 (D). CpG 1 and 2 were sequenced by SP1, CpG 3 and 4 were sequenced by SP2.

4.3.8 Methylation differences between twin pairs

Lebastchi *et al*. and Husseiny *et al*. analysed samples from newly diagnosed patients (Lebastchi *et al.*, 2013; Husseiny *et al.*, 2014). At diagnosis, most of the β-cells are destroyed, however patients with disease duration for over 50 years, still show some signs of insulin production (Keenan *et al.*, 2010). Here, samples were taken from five twin pairs where the index twin was diagnosed less than two years prior. The methylation levels from these twin pairs were compared to the remaining twins with long-standing diabetes at each CpG site to see if insulin DNA could still be detected in those with disease duration of more than two years from sampling (Figure 4.15).

There were no significant differences between the groups at CpG 1 (Figure 4.15A), CpG 2 (Figure 4.15B), CpG 3 (Figure 4.15C) and CpG 4 (Figure 4.15D). Although at CpG 1, it was noticeable that there was hypomethylation in samples from the twins with shorter disease duration, compared to individuals with long standing diabetes and both groups of healthy co-twins ($p < 0.0536$).

Intra- and inter-pair methylation differences were also analysed (Figure 4.16). The inter-pair differences were expected to be greater in diabetic twins with shorter disease duration than those with long-standing diabetes. This was observed at CpG 1 ($p = 0.0398$), but not at CpG 2 ($p = 0.8959$), CpG 3 ($p = 0.3368$) and CpG 4 ($p = 0.9083$). There were also no significant intra-pair differences.

Intra- and inter-pair methylation differences were analysed in groups separated by age at diagnosis (Figure 4.17). This analysis was performed to see whether the more aggressive form of diabetes in younger patients made a difference in the level of unmethylated insulin DNA compared to adult-onset diabetics. There were no significant intra- and inter-pair differences between the twin pairs.

In summary, hypomethylation in the diabetic twin was observed at CpG 1 only due to disease duration, rather than age at diagnosis.

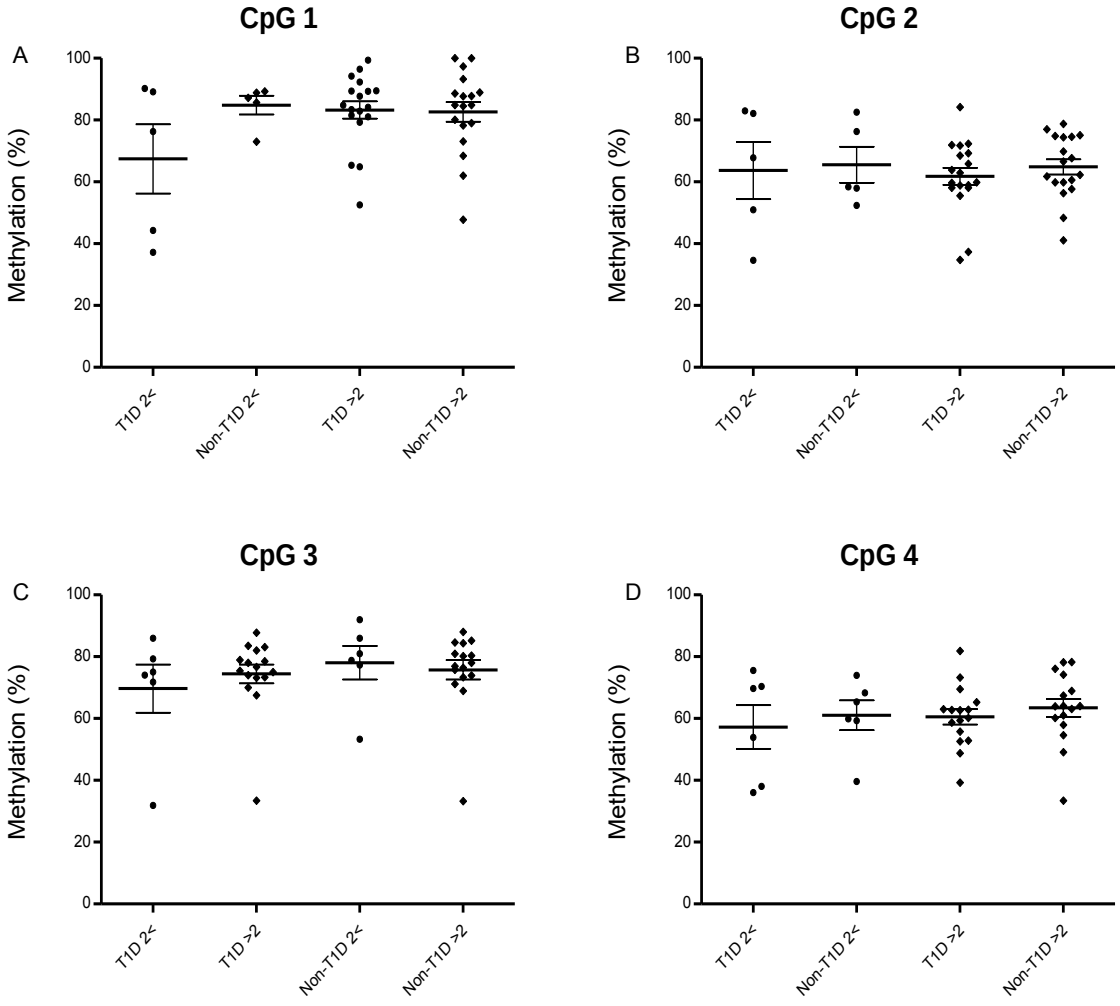


Figure 4.15. Methylation levels of each individual sorted by disease duration. The two different groups were separated into disease duration at sampling at less than two years and long-standing diabetes. Methylation levels were plotted for CpG 1 (A), CpG 2, (B), CpG 3 (C) and CpG 4(D). Data presented as mean± SEM.

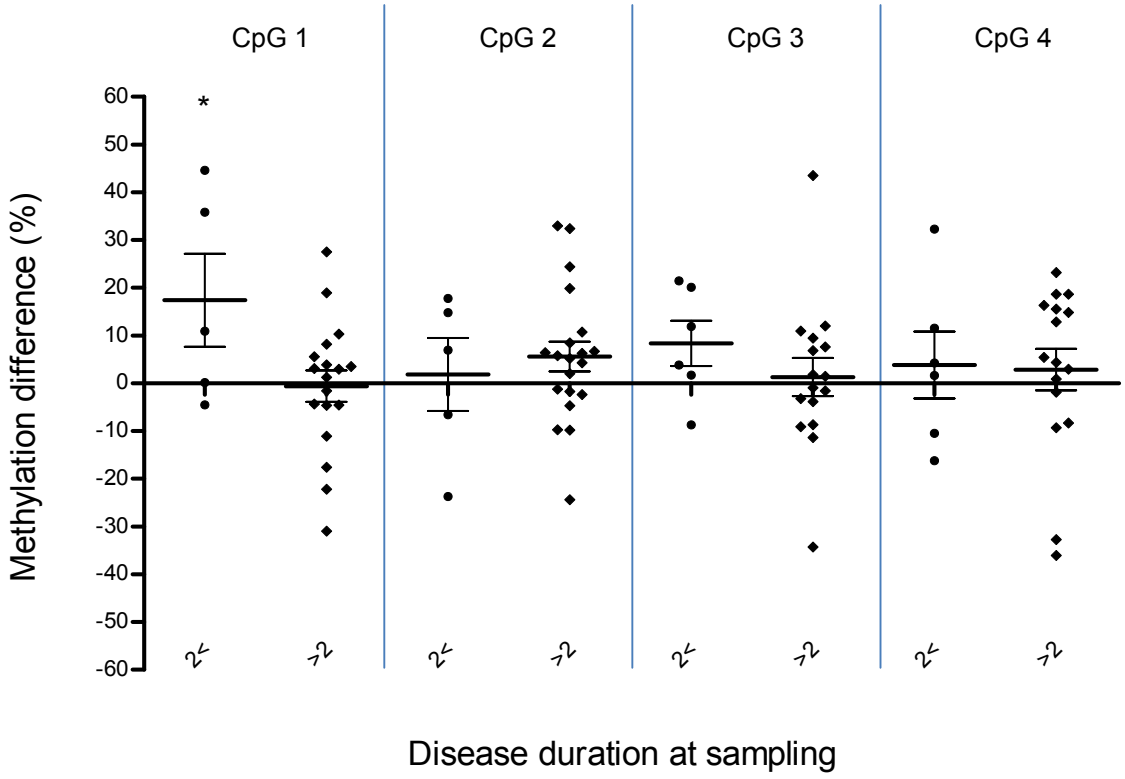


Figure 4.16. Methylation differences between twin pairs sorted by disease duration. The methylation differences (non-diabetic – diabetic) were compared between the two groups at each CpG site. There was one significant result between the newly-diagnosed group and long-standing diabetes group at CpG 1 (*p = 0.0398), whereas there were no significant intra- or inter-pair differences at the other CpG sites. Data presented as mean± SEM.

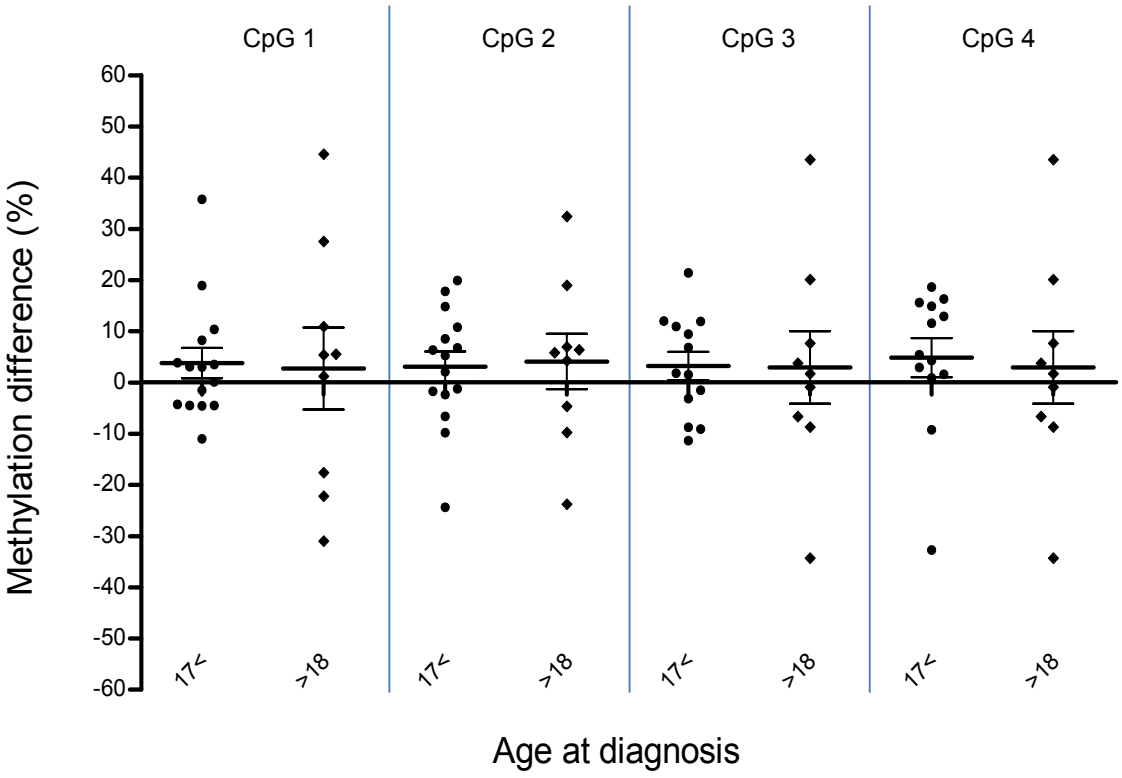


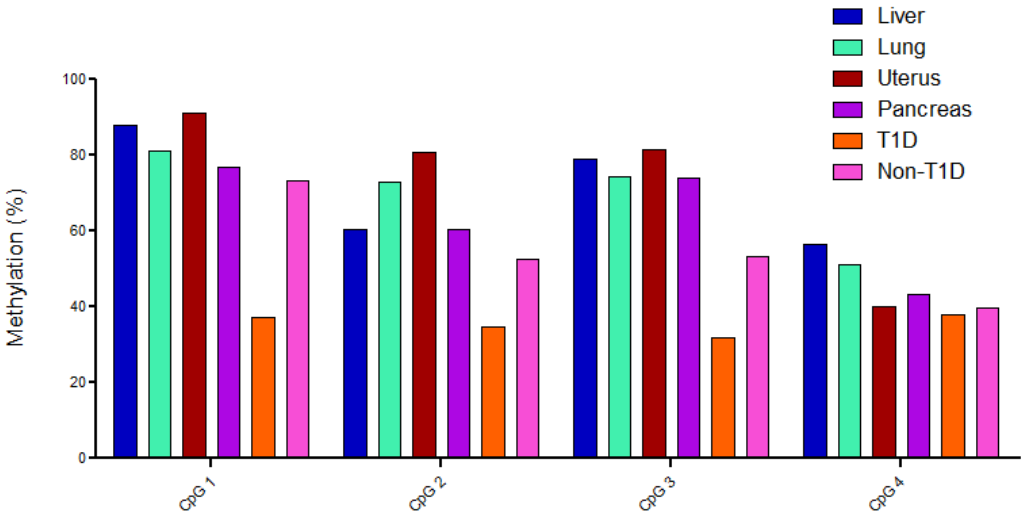
Figure 4.17. Methylation differences between twin pairs sorted by age at diagnosis. The methylation differences (non-diabetic – diabetic) were compared between childhood-onset diabetes to adult-onset diabetes at each CpG site. There were no significant intra- or inter-pair methylation differences. Data presented as mean± SEM.

4.3.9 Tissue-specific unmethylation in the β-cells

In the body, many cells can break down, releasing DNA into the circulation. To assess whether the unmethylated CpG sites at -206 and -135 were tissue-specific, DNA from liver, lung, uterus and pancreas were commercially obtained and treated with sodium bisulfite. The samples were then amplified and sequenced.

Methylation levels were similar in liver, lung, uterus, pancreas as well as in the non-diabetic twin at all CpG sites (Figure 4.18A). The T1D sample was from a newly-diagnosed twin and the methylation levels in the diabetic twin sample was much lower than the other tissues at CpG sites 1 (-206), 2 (-180), and 3 (-135). There was also a noticeable difference between the diabetic and non-diabetic twin at CpG sites 1, 2 and 3 but not 4. The methylation levels of the diabetic samples at CpG 1, 2, 3 and 4 were 37.22%, 34.62%, 31.91% and 38.00% respectively. However, when the newly-diagnosed twin pair was swapped with a pair whose index twin had long-standing diabetes, methylation levels in all samples were similar across all CpG sites (Figure 4.18B). The hypomethylation in the newly-diagnosed T1D sample was also evident in a heatmap analysis (Figure 4.19A and B).

A



B

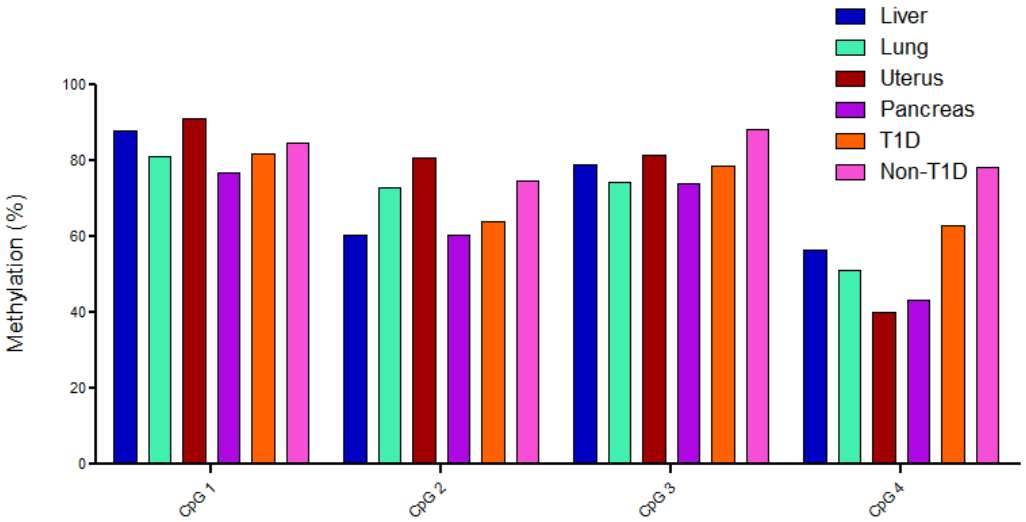


Figure 4.18. Methylation differences between a newly-diagnosed diabetic compared to one with long-standing diabetes. Liver, lung, uterus and pancreas DNA were commercially obtained and were treated with sodium bisulfite. At CpG 1, 2 and 3, methylation level of the newly-diagnosed diabetic twin was much lower than the other tissues and non-diabetic twin (A). However, in long-standing diabetes, methylation levels were similar to the other tissues across all CpG sites (B).

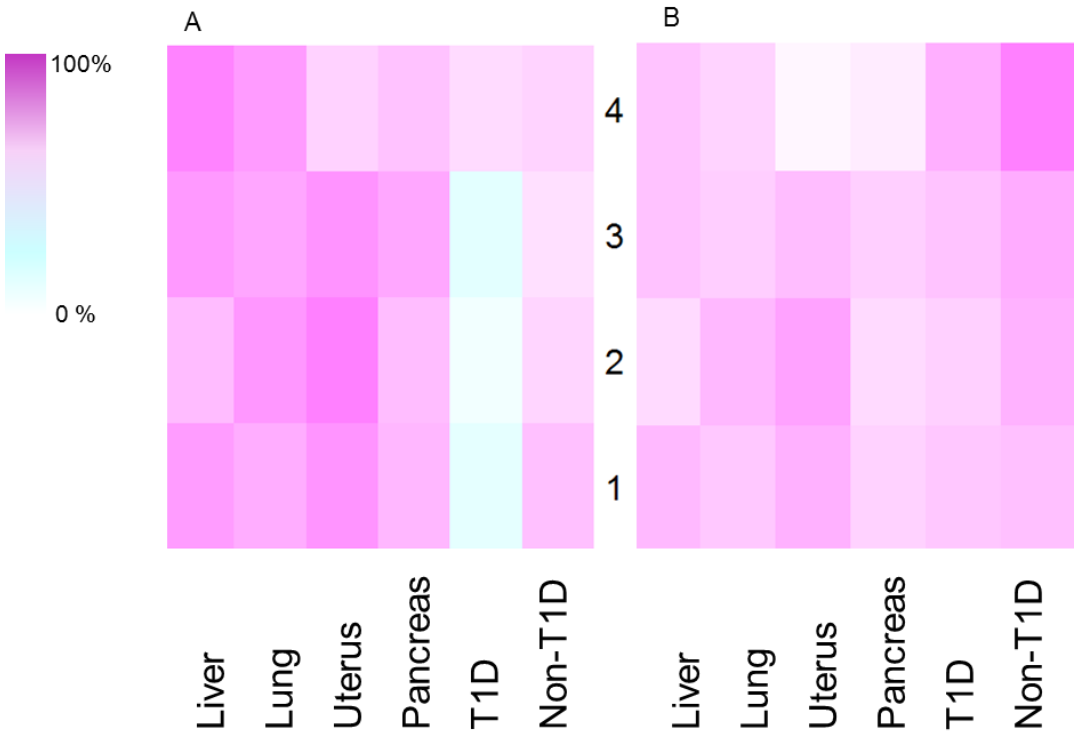


Figure 4.19. Methylation status in different tissues. Heatmaps of the methylation levels from each tissue. Hypomethylation was observed in the sample from the newly-diagnosed diabetic (A), compared to the sample from the patient with long-standing diabetes (B). All other tissues had similar methylation levels.

4.4 Discussion

4.4.1 Primers did not discriminate between methylated and unmethylated CpG sites in exon 2

This study investigated the methylation status in the insulin gene. Insulin is expressed mainly in β-cells and in T1D, these cells are destroyed, releasing DNA into the circulation. Free circulating DNA has been investigated as a biomarker for disease (Swarup and Rajeswari, 2007; Holdenrieder *et al.*, 2008). It is a non-invasive, easy source of DNA to obtain and although the concentration of DNA in serum or plasma is low, sensitive assays such as qPCR can detect it.

This method for monitoring β-cell death was previously reported (Akirav *et al.*, 2011; Lebastchi *et al.*, 2013) and had described a working nested-qPCR assay in detecting methylated and unmethylated CpG sites that were tissue-specific to β-cells. Originally this method was performed using serum from mice and then trialled using human serum. A few changes were made to the method the authors had described. For example, they had included a gel purification step between the PCR and qPCR steps. This was replaced with a simple 1/1000 dilution of the first PCR product as the resulting amplification was the same. 2 µL of the diluted product was then used as a template for the qPCR reaction. This saved time and the cost of the commercial kit required for the gel purification step. Also, DNA recovered from the spin columns of the purification step may affect the total concentration and therefore skewing the number of unmethylated to methylated DNA. The main issue from this assay was that there was no differentiation between detecting unmethylated from methylated DNA in serum with the pre-designed primer sets. This may be due to the assay detecting methylated DNA from other tissues. The authors had only measured unmethylated insulin DNA in one other tissue, the kidney, which was found to be mainly methylated. However, there may be other tissues that may have the same trend in methylation as the β-cells. The non-specificity of the primers was also reported by another group (Husseiny *et al.*, 2012).

It also has been noted that the calculation Akirav *et al*. performed did not account for the efficiency of the assay (Husseiny *et al.*, 2012). When the amplification efficiency was performed for all primer sets, they came up to less than 100%, therefore the Cq values are not completely reliable as they products are not amplifying at 100%.

4.4.2 Real-time PCR detected unmethylated insulin DNA but primer dimer issues occurred

After the experiment described in Section 4.3.1 was unsuccessful, a different approach was taken. Husseiny *et al*. described a method that detected only specific unmethylated CpG sites that were tissue-specific to β-cells, also using qPCR (Husseiny *et al.*, 2014). As with Lebastchi *et al*., this method was first performed using serum from mice, and then with human patient serum. Husseiny *et al*. had analysed unmethylated insulin DNA in eight other tissues, including liver, lung and spleen, and reported that at CpG sites -206 and -135, were tissue-specific unmethylated compared to the other tissues. Moving forward with this, a nested-MSP was designed to detect these CpG sites with primers discriminating between methylated and unmethylated CpG sites. The first step required the BSP primer set to amplify the total amplifiable region for analysis. During optimisation of all primer sets, primer dimers appeared with BSP. This was an issue as any signal that was detected using the BPS primer, could have been from the primer dimer and therefore was not able to sufficiently quantitate the methylation levels from a single sample.

The issue with primer dimers did not resolve itself after numerous optimisation steps, therefore the Taqman technology was used. Taqman is highly specific, hence it was expected that primer dimers would not appear during this assay. However, when this assay was performed with a healthy control sample, the signal for MSP2 was lower than the signal for BSP. This was unexpected as firstly, the result was the opposite of what was seen when using SYBR Green and secondly, as the BSP primer did not discriminate between unmethylated and methylated CpG sites, in theory it should have amplified more DNA than the MSP2 primers as the MSP2 primers would amplify DNA with unmethylated CpG sites.

4.4.3 Pyrosequencing revealed hypomethylation at CpG site -206

As the qPCR assays were not able to sufficiently detect and discriminate between methylated and unmethylated templates, pyrosequencing was performed to determine the methylation levels of each sample. Pyrosequencing assesses DNA methylation by sequencing-by-synthesis, providing accurate and reliable results. Two sequencing primers were used to cover the two CpG sites of interest, -206 and -135.

There were a few interesting points from the data that was presented in Figure 4.13. First, the plot showed that there were no significant differences between the diabetic and their healthy co-twin at each CpG site. This could account for why no differences in methylation were seen between a MZ twin pair discordant for T1D in Section 4.3.1, Figure 4.2. Second, methylation levels were higher at CpG sites 1 and 3 compared to 2 and 4. This was opposite to what was expected. This is because it has been reported that CpG 1 (-206) and 3 (-135) were unmethylated in a tissue-specific way, compared to CpG sites 2 (-180) and 4 (-102), which were mainly unmethylated across all nine tissues (Husseiny *et al.*, 2014). Hence, an explanation for the general hypomethylation seen at CpG 2 and 4 could be that DNA from other tissues in the circulation that the assay was detecting was contributing to the overall hypomethylation. Third, a higher methylation state at CpG 1 and 3 could be down to the β-cells. In the papers published by Akirav *et al*. and Husseiny *et al*., CpG 1 and 3 in the clones they had produced, were not completely unmethylated (Akirav *et al.*, 2011; Husseiny *et al.*, 2014). Therefore, even though the unmethylated CpG sites were associated with β-cells, it is possible that the methylated DNA could be released by damaged β-cells. A way to potentially find out how much DNA from the tissues is contributing to the differing methylation levels is to analyse samples from patients who had undergone total pancreatectomies. These patients will have no β-cells, therefore it will be a good indicator whether or not DNA methylation that was detected originated from β-cells or not.

The twin pairs were then separated by disease duration. The groups were split into those who were diagnosed less than two years from sampling and the remaining twin pairs, who were considered to have long-standing diabetes. There was a significant methylation difference between the newly-diagnosed group and long-standing diabetes at CpG 1 ($p = 0.0398$) (Figure 4.16). However, no significant intra- or inter-pair differences were detected at the other CpG sites. In newly-diagnosed T1D patients, most of the β-cells may still be undergoing apoptosis and therefore still releasing DNA, and as CpG 1 has been reported to be unmethylated in β-cells, the significant methylation difference between newly-diagnosed and long-standing diabetes was expected. However, it does not explain why then there were no differences between the diabetic and non-diabetic twin pair at the other CpG sites. There were also no differences between groups sorted by age at diagnosis, suggesting that hypomethylation in the diabetic twin observed at CpG 1 was due to disease duration, rather than age at diagnosis.

4.4.4 Insulin promoter hypomethylation was a feature in newly-diagnosed T1D sample compared to other tissues

In this study, DNA from other tissues was sequenced (Figure 4.18). There were no significant methylation differences between liver, lung, uterus and pancreas but there was noticeable hypomethylation in the sample from the newly-diagnosed diabetic. Higher methylation levels were seen in those with long-standing diabetes and this may be due to the fact that most of the β-cells are destroyed and the half-life of insulin is approximately 5 minutes (Duckworth *et al.*, 1998).

Although in the several clones Akirav *et al*. and Husseiny *et al*., had produced, were not completely unmethylated, methylation at the CpG sites may not affect insulin transcription (Kuroda *et al.*, 2009). In mice, the insulin promoter region only has three CpG sites. Kuroda *et al*. assessed individual methylation events at each CpG site. They reported that of the three sites, only one supressed insulin promoter activity by approximately 50%. And complete methylation of the human insulin gene promoter suppressed gene activity by 85%. However, the authors suggested that β-cell specific demethylation was involved in the differentiation of β-cells rather than being involved in disease process. They observed that when culturing β-cells with normal glucose levels, the insulin promoter was unmethylated, suggesting that either the insulin gene was demethylated in the embryo and becomes preferentially methylated in all cells but the β-cells, or the gene was methylated in the embryo and becomes specifically demethylated in β-cells. They tested this hypothesis in the mice models and found in early differentiation steps in mouse embryonic stem cells, there was methylation. Complete demethylation was observed at the late step of differentiation of the β-cells. Therefore supporting the hypothesis that certain CpG sites in the insulin promoter region is unmethylated in a tissue-specific manner.

4.5 Conclusion

In conclusion, early qPCR experiments detecting methylated and unmethylated CpG sites in exon 2 of the insulin gene were unsuccessful due to non-specificity of the primers. The attention then turned to the insulin promoter region where using SYBR Green and Taqman detected unmethylated CpG sites, however issues with primer dimer formation occurred. This made it difficult to determine whether the signal detected was from the amplified DNA of interest or from the primer dimers.

Pyrosequencing was performed to acquire an accurate reading of bisulfite converted DNA in order to identify potential differences in methylation between MZ twin pairs discordant for T1D. Across all four CpG sites interrogated, there was no significant difference seen between the twin pairs, however when comparing a cohort of newly-diagnosed patients with those with long-standing diabetes, there was a significant difference between the two groups. In newly-diagnosed patients, there was hypomethylation in newly-diagnosed diabetics at CpG site -206 in the insulin promoter compared to their healthy co-twin and in other tissues such as liver and lung. This suggests that at time at diagnosis, it is possible to detect DNA methylation from β-cells, therefore it is essential that sampling is performed within one or two years of diagnosis for monitoring of β-cell death.

Chapter 5

General discussion

5 General discussion

Epigenetic modifications play a major role in the regulation of gene expression (Bird, 2007; Irizarry *et al.*, 2009). DNA methylation is the best characterised epigenetic mark and has been associated with different diseases, most notably cancer (Laird, 2003; Feinberg, 2007). Different environmental agents can affect DNA methylation including smoking (Besingi and Johansson, 2014) and dietary sources such as folic acid and methionine (Cooney *et al.*, 2002). T1D is a disease that is characterised by the destruction of insulin producing β-cells in the pancreas and accounts for 5-10% of all diabetes cases worldwide (American Diabetes, 2009). T1D is the result of the interaction of genetic and environmental factors (Atkinson *et al.*, 2014) and it was suggested that epigenetics could play a part in the aetiology of T1D. In T1D, it has been reported that there were differences in methylation between MZ twin pairs discordant for the disease (Rakyan *et al.*, 2011a). Studying epigenetics and disease in twin pairs is useful in removing genetic confounding factors (Bell and Saffery, 2012) and there have been differences between MZ twin pairs in disease (Stefan *et al.*, 2013) and an effect has been seen between twins of different ages (Fraga *et al.*, 2005).

In this thesis, the aim was to investigate the role of DNA methylation in T1D by generating whole-genome-wide methylation profiles from $CD14^+CD16^-$ monocytes, $CD4^+$ T cells, $CD19^+$ B cells and buccal cells. Different DNA profiling technologies were used. Also, an assay was developed to detect differences in DNA methylation in circulating DNA. From these studies, it was further established that DNA methylation marks are stable and that these differences were cell type specific. Also, DNA methylation differences could be detected in circulating DNA, which could be used as a potential biomarker.

5.1 DMPs identified in the twin pairs were cell type-specific

Previously, DNA methylation differences between MZ twin pairs discordant for T1D have not been studied in $CD14^+CD16^-$ monocytes, $CD4^+$ T cells, $CD19^+$ B cells and buccal samples in a single study. The data presented in this thesis showed cell type-specific DNA methylation differences between MZ twin pairs in the immune cells. The methylation differences were detected in nine of the 16 twin pairs that were sampled in a previous study carried out years before, therefore indicating that DNA methylation is a stable marker.

Cell composition is important as each cell type has its own methylation profile (Reinius *et al.*, 2012; Lam *et al.*, 2012). Although it is easier to ascertain PBMC samples, profiling specific cell types is crucial. Ideally, fresh samples from patients would be obtained and processed immediately, however this may not always be possible and studies have used cell lines to assess methylation (Stefan *et al.*, 2013). Stefan *et al*. had used EBV transformed lymphoblastoid cell lines for DNA methylation profiling. However, DNA methylation profiles differ in leukocytes from peripheral blood to lymphoblastoid cell lines as the addition of EBV to immortalize human cells induces large-scale hypomethylation (Brennan *et al.*, 2009; Hansen *et al.*, 2014), therefore introducing bias. Another readily available type of sample is formalin-fixed paraffin-embedded (FFPE) tissue. A protocol has been developed to study archival FFPE tissue on the Illumina platform (Thirlwell *et al.*, 2010). Although laboratories may not have access to fresh peripheral blood to study methylation, other types of tissue can be profiled. However, caution must be taken when analysing cell- or tissue-specific methylation.

Hundreds of genetic variants identified by GWAS have been associated with complex diseases and deep sequencing can reveal rare variants that can contribute to disease pathogenesis (Hunt *et al.*, 2013). However, most of the variants confer small increments in risk and missing heritability may account for the remaining risk (Manolio *et al.*, 2009). The small DNA methylation differences shown in this thesis can contribute to missing heritability. It has been suggested that epigenetic variation can be an explanation for missing heritability these modifications are stable and passed on through generations, therefore contributing to the risk of developing a disease (Slatkin, 2009; Koch, 2014). However, small differences in methylation between the twin pairs do not compare to differences found in alcohol consumption (~20%) (Zhu *et al.*, 2012) and smoking (~12%) (Breitling *et al.*, 2011). Small significant differences between the twin pairs in this study were a feature in the power calculation that was performed. It was then suggested to ascertain at least 50 twin pairs for sufficient study power of over 90%.

5.2 The effects of ageing on methylation

DNA methylation has been reported to have an inverse correlation with age in humans (Langevin *et al.*, 2011; Steves *et al.*, 2012) and in rats (Thompson *et al.*, 2010). This may be due to the cellular composition of blood being altered due to ageing, therefore presenting with a different methylation profile (Jaffe and Irizarry,

2014) and this is referred to as 'epigenetic drift' (Teschendorff *et al.*, 2013b; Fraga and Esteller, 2007). This was supported by a study which reported over 10% methylation change over time in the same individual over 11 years (Bjornsson *et al.*, 2008). Intra-pair DNA methylation differences were described in childhood-onset diabetic patients (Rakyan *et al.*, 2011a). This present study consisted of childhood-onset and adult-onset diabetics, and it was expected that there would be larger methylation differences between adult-onset diabetics and their healthy co-twin due to age. However, this was not the case. Conversely, the more significant methylation differences came from twin pairs who were diagnosed with T1D at a younger age. This may be due to two things. First, there was a genuine effect or secondly, the effect was seen as those twins were the same pairs from the preliminary study where they analysed $CD14^+$ monocytes from childhood-onset diabetic twin pairs and the same T1D-MVPs calls were identified. More twin pairs would then be needed to shed some light on this issue.

The age of a human sample can now be predicted (Horvath, 2013) and it is important to have age-matched controls considering the change to an individual's methylome due to age when studying different diseases. Age-related DNA methylation changes have also been seen in a young cohort (Martino *et al.*, 2011). Mononuclear cells were collected from new-borns and were subsequently sampled at ages 1, 2.5 and 5 years. There was a clear distinction in methylation between the age groups and after functional analyses, genes associated with methylation changes were with cell surface receptor and signal transduction events. An ideal study design may be to obtain samples from birth in the form of Guthrie cards or cord blood and follow up with samples from childhood through to adulthood. Guthrie cards gives a snapshot of what might have happened in utero (El Hajj *et al.*, 2013) already showing DNA methylation differences in genes at birth (Lee *et al.*, 2012a) and has been reported to be suitable for genome-wide DNA methylation profiling (Joo *et al.*, 2013; Aberg *et al.*, 2013). Limitations to studying these blood spots include issues with handling and storage, contamination, DNA degradation and they are difficult to obtain due to ethics (Ramagopalan and Rakyan, 2013). However, generating DNA methylation profiles from birth will provide information on epigenomic variation in different human diseases.

5.3 Illumina450K versus bisulfite sequencing

In Chapter 3, two different DNA methylation profiling technologies were used. The Illumina450K BeadChip and BS-seq. Illumina450K interrogates over 485,000 CpG sites and contains two chemistry technologies: 135,000 probes from Infinium I array, 350,000 probes from Infinium II (Sandoval *et al.*, 2011). One advantage of this array is that the probes can be designed for CpG-poor regions, such as CpG island shores and shelves (Touleimat and Tost, 2012; Dedeurwaerder *et al.*, 2011). A limitation of Illumina450K is that it does not generate a whole genome-wide methylation profile, however this is one of the advantages of performing a BS-seq analysis.

As seen in Chapter 3, BS-seq resulted in more hypomethylated sites compared to Illumina450K. This was due to noise therefore requiring deeper coverage which costs more money. Running both platforms is ideal for comprehensive coverage of the methylome, however due to costing, one could be used for the main experiment and the other for validation. Limitations in both the Illumina450K array and BS-seq include incomplete conversion of unmethylated cytosines. Incomplete conversion will lead to false positives in methylation. Another limitation is the degradation of DNA during treatment with sodium bisulfite. Up to 90% of DNA is degraded during the process due to the high bisulfite concentrations, elevated temperatures and long incubation times (Tanaka and Okamoto, 2007). In both the Illumina450K array and BS-seq, bisulfite treatment cannot distinguish between 5-mC and 5-hmC (Krueger *et al.*, 2012). Recently a new method has been proposed to detect this difference (Stewart *et al.*, 2014). The authors had adapted a bisulfite-based 5-mC profiling technology for 5-hmC detection. They combined oxidative bisulfite (oxBS) chemistry with the Illumina450K BeadChip and have shown that the oxBS-450K protocol produced results that were highly reproducible compared to BS-seq. There are several other profiling technologies currently used by investigators (Bock, 2012; Liu *et al.*, 2012; Lister and Ecker, 2009). On occasion, two different profiling methods can be combined for more comprehensive methylome coverage (Harris *et al.*, 2010). These technologies are described in Section 1.8.3.1.

As well as different profiling technologies, there are many pipelines that specialise in analysing DNA methylation data such as Minfi (Aryee *et al.*, 2014), WateRmelon (Pidsley *et al.*, 2013) and RnBeads (Assenov *et al.*, 2014). Each has its own uses (Wilhelm-Benartzi *et al.*, 2013; Morris and Beck, 2014). In this study, ChAMP was

performed in R. One other disadvantage of Illumina450K is the bias seen across the two different probes, however different R packages work to correct that bias (Teschendorff *et al.*, 2013a). With profiling technologies costs lowering, there is a large number of DNA methylation data. A database that incorporates Illumina450K data from different methylation studies called Marmal-aid (Lowe and Rakyan, 2013). Marmal-aid currently holds DNA methylation data on 14,586 samples (true as of 29/12/14) and is continuing to grow in size. This database enables an investigator to reprocess data and visualise the methylation state in different tissues.

The use of certain profiling platforms and analysis methods is important, however, a confounding factor is different isolation methods, which affect analysis and therefore future functional studies. In a study comparing human monocyte functionality, different methods of cell isolation were tested. Anti-CD14 magnetic microbeads, non-monocyte depletion by antibody-conjugated magnetic microbeads, RosetteSep antibody cocktail and the classical adherence protocol were performed and it was reported that positive selection with magnetic microbeads gave the best results in term of purity and cell functionality (Zhou *et al.*, 2012). Global DNA methylation is also affected by different DNA extraction methods (Soriano-Tarraga *et al.*, 2013) and storage of samples (Mallone *et al.*, 2011), therefore it is important to batch samples for analysis.

5.4 DNA methylation differences detected in serum

In diabetes, β-cells are invaded by immune cells leading to the destruction of these insulin secreting cells. Current methods of monitoring β-cell death are measuring fasting C-peptide levels (Wang *et al.*, 2012) and PPP1R1A (Jiang *et al.*, 2013). Chapter 4 described the use of nested methylation-specific PCR and pyrosequencing to determine levels of unmethylated insulin DNA in MZ twin pairs discordant for T1D.

The qPCR experiments with SYBR Green and Taqman were not able to successfully detect unmethylated and methylated CpG sites, therefore pyrosequencing was performed. Pyrosequencing resulted in a significant difference in methylation between twin pairs discordant for T1D where the diabetic was sampled less than two years from diagnosis. A limitation to this study was perhaps the newly-diagnosed patients. The sample size was small (5 vs 19 long-standing diabetics) and this difference would not be detected in those with long-standing

diabetes. This may be that newly diagnosed patients still have β-cells compared to those with long-standing diabetes. Therefore, to determine how early this methylation difference can be detected, individuals who are at risk of developing T1D, that is those who have one or more diabetes-associated autoantibodies, should be studied. This would be useful to understand how quickly the autoimmune process progresses during the break down of the β-cells.

There are several studies investigating different immunotherapies to preserve or regain β-cell regeneration and function. Lebastchi *et al.* reported preservation of β-cell function in patients who received teplizumab, an anti-CD3 monoclonal antibody, compared to the placebo (Lebastchi *et al.*, 2013). Another group administered lysine deacetylase inhibitors which saw a decrease in islet inflammation and increased numbers of Tregs in NOD mice (Christensen *et al.*, 2014). Deciphering the β-cell epigenome in diabetes can have translational implications (Johnson and Evans-Molina, 2014; Bramswig and Kaestner, 2012), leading to efforts to prevent, diagnose and monitor diabetes.

5.5 Comparison of the different techniques to detect unmethylated CpG sites in circulating DNA

In Chapter 4, patient serum was used to detect DNA methylation differences between twin pairs. Plasma also contains free circulating DNA, however it has been shown that higher amounts of DNA can be extracted from serum compared to plasma (Holdenrieder *et al.*, 2005; Umetani *et al.*, 2006). Either way, the assays performed in Chapter 4 were able to detect low levels of bisulfite converted DNA. At first, SYBR Green was used in qPCR as described by other groups studying methylation in T1D (Husseiny *et al.*, 2014; Lebastchi *et al.*, 2013), but also in cancer (Anker *et al.*, 2001). SYBR Green is also inexpensive and easier to design than Taqman. However, as described in Section 4.3.2, primer dimers occurred due to the non-specificity nature of the assay and therefore could not be used. The use of the Taqman assay, although highly specific, is more costly. Another advantage of using Taqman is that it allows multiplexing to be performed, however, if interrogating more than two CpG sites, more probes will have to be designed, therefore raising the cost of the experiment. Pyrosequencing combines PCR and sequencing-by-synthesis techniques to provide accurate CpG methylation measurement. Advantages of performing pyrosequencing are the cost for this assay is lower than the Sanger method, hundreds of thousands of sequence reads can be obtained in a single run

and is very accurate (Siqueira *et al.*, 2012). A limitation in performing pyrosequencing, is that the assay can only sequence short DNA sequences. Hence, in this study two sequencing primers were designed to assess the methylation levels of two CpG sites. A limitation in this study was the small sample size, particularly when comparing the newly-diagnosed twins to long-standing diabetes twins (5 to 19 pairs respectively). Although there was a significant result in Figure 4.16, a power calculation should also be performed. Another limitation was primer design, this applied to all assays. The primers were checked for self-looping, primer dimer hybridization and cross-hybridization, however, primer dimer issues can still occur.

qPCR can be performed in a standard diagnostic laboratory in a hospital setting and therefore would be a useful way to apply this method for monitoring β-cell death. It is also inexpensive and simple to perform compared to pyrosequencing, as not every laboratory will have a pyrosequencer. Another technique based on the same principle is Droplet Digital™ PCR (ddPCR) (Usmani-Brown *et al.*, 2014). ddPCR involves the sample which is dispersed into droplets that acts as individual PCR reactions. The sensitivity of the assay is higher than qPCR and reduces contamination as there is no need for nested PCR. The ddPCR system would also be easily obtainable in a diagnostic laboratory.

5.6 Biomarkers of β-cell death in diabetes

In T1D, autoantibodies against GAD65, IA-2 and ZnT8 are established biomarkers for the disease (Dang *et al.*, 2011; Leslie *et al.*, 2001). As it is still unknown whether T1D-MVPs are the outcome of autoantibody presence or is involved in the seroconversion, they could potentially become biomarkers for T1D. DNA methylation differences seen between the twin pairs in this thesis indicate that methylation marks are stable and can be detected in the circulation. Further studies will have to be performed to observe whether the difference in unmethylated insulin DNA in diabetic patients compared to healthy controls, is significant enough to be considered as a diagnostic or monitoring tool for β-cell death. Nevertheless, other potential biomarkers have been studied and reported in diabetes.

Recently, preproinsulin (PPI) has been proposed as a potential biomarker to detect β-cell death (Fisher *et al.*, 2013). PPI is a precursor molecule of insulin and following sequential cleavage of PPI and proinsulin, insulin is synthesised in the β-cells. The group developed an assay to detect the methylation status of circulating PPI DNA in

mice. Using dual fluorescent-probe multiplex PCR, they showed a significant increase in demethylation prior to onset of hyperglycemia. The authors also reported a decrease in demethylation following diabetes development, likely reflecting the absence of further β-cell death. Therefore this assay could be used in detect β-cell death in newly-diagnosed patients. Key miRNAs can also be detected in circulation, therefore demonstrating the potential to become predictive biomarkers for disease (Nielsen *et al.*, 2012; Klein *et al.*, 2013). miRNAs are small non-coding RNA molecules that are involved in post-translational regulation of gene expression (Tomankova *et al.*, 2011). Nielsen *et al.* compared expression levels of sera miRNA related to β-cell function and glycaemic control from newly-diagnosed T1D children and age-matched healthy controls. They reported several miRNAs that were in upregulated in the T1D patients such as miR-27a, miR-29a and miR-210. In particular, miR-25 was significantly negatively associated with residual β-cell function ($p = 0.0037$). This study suggested that miR-25 may be tissue-specific for glycaemic control after T1D diagnosis therefore may be a useful biomarker for the disease. Epigenetic alterations in disease have a crucial role for the clinics. These modifications can pave the way for personalised medicine including for future diagnosis, prognosis and prediction of response to therapies (Heyn and Esteller, 2012; Herold *et al.*, 2013).

Chapter 6

Conclusion and further work

6 Further work and conclusion

6.1 Further work

The data from this thesis provided new evidence on DNA methylation differences in MZ twin pairs discordant for T1D. More twin pairs will be ascertained as the power analysis suggested recruiting 50 MZ twin pairs to obtain 90% power. Once the DNA methylation signatures have been identified in the discovery cohort, it is important to replicate and validate the findings. As the DMPs were cell-specific, these DMPs may be found in whole blood. Sorting cells from blood is a costly and time-consuming process, therefore if the signatures can be detected in whole blood, sampling and processing of samples would be easier. To perform the replication experiment, whole blood samples from T1D singletons and control singletons will be analysed. There are several methods in which the methylation signatures can be detected, one of which is MethyLight. MethyLight is based on Taqman technology where several CpG sites can be interrogated at the same time through multiplexing (Eads *et al.*, 2000). The assay requires small DNA concentrations, approximately 50 ng, and is relatively simple to perform. In addition to the 24 twin pairs profiled on the Infinium450K BeadChip, eight $CD4^+$ lymphocyte samples from four MZ twin pairs were analysed using BS-seq. Validation will involve $CD4^+$ lymphocytes from the 24 pairs and profiling them on the Fluidigm platform. The Fluidigm platform enables the investigator to study DNA methylation in a single cell. Once the signatures have been replicated and validated, it would be useful to understand the functional outcomes of these DNA methylation differences at a molecular level, how it affects the chromatin structure and its impact on gene expression.

Rakyan *et al.* had analysed individuals who were autoantibody positive but did not develop T1D (Rakyan *et al.*, 2011a). Methylation differences were detected when comparing these high risk individuals with controls, leading to the question whether or not these methylation differences arise before or after seroconversion of diabetes-associated autoantibodies. If the DMPs appeared before the presence of autoantibodies, this would suggest that they could cause disease and not become the result of secondary autoantibody-associated immune process. If the DMPs were detected before autoantibody positivity, then these marks could potentially be detected at birth. Although there was a small correlation found in cord blood, follow up samples from the 99 children are currently being collected to see if these signatures are detected much later on.

Although detecting methylation differences was difficult through the use of qPCR, differences were found between newly diagnosed twin pairs compared to twins who had long-standing diabetes through pyrosequencing. The results from pyrosequencing set a platform for further studies into β-cell death. In both qPCR and pyrosequencing assays, it was possible that the DNA amplified, may have come from other tissues. An experiment analysing samples from patients who have had total pancreactomies may shed some light on the source of circulating DNA.

6.2 Conclusion

This thesis has investigated DNA methylation in T1D. In a preliminary study, it was shown that DNA methylation antedated T1D and differences were identified between MZ twin pairs discordant for T1D. Following on from this, the present study investigated whole genome-wide DNA methylation profiles using a more extensive array (Illumina450K) in a broader range of cells, $CD14^+CD16^-$ monocytes, $CD4^+$ T cells, $CD19^+$ B cells and buccal cells from a larger cohort of MZ twin pairs discordant for T1D.

Evidence presented here has shown that DNA methylation marks are stable, at least for three years. Although DNA methylation differences between the twin pairs were small, DMPs identified were cell type specific. DMRs were also found to overlap $CD4^+$ enhancer sites through the use of BS-seq, and therefore further investigation into chromatin remodelling and disease should be performed. However, it was found that even larger number of twins are needed to achieve sufficient power (>85%). Further replication and validation studies are required in turn to confirm disease-associated epigenetic markers and could be of diagnostic or therapeutic value. Detection of DNA methylation differences in serum could present a novel way to monitor β-cell death and help in the assessment of immunotherapy.

There is evidence to suggest that DNA methylation has a role in disease phenotype, whether it is a cause effect or the result of disease remains to be investigated. Integration of GWAS and EWAS data can pave the way to better understand complex autoimmune diseases such as T1D.

Chapter 7

References

7 References

Aberg KA, Xie LY, Nerella S, Copeland WE, Costello EJ and van den Oord EJ (2013). High quality methylome-wide investigations through next-generation sequencing of DNA from a single archived dry blood spot. *Epigenetics.* 8(5):542-7.

Absher DM, Li X, Waite LL, Gibson A, Roberts K, Edberg J, Chatham WW and Kimberly RP (2013). Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. *PLoS Genet.* 9(8):e1003678.

Accomando WP, Wiencke JK, Houseman EA, Nelson HH and Kelsey KT (2014). Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol.* 15(3):R50.

Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, Smith AV and Gudnason V (2012). Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS One.* 7(10):e46705.

Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, Dahl F, Dermitzakis ET, Enver T, Esteller M, Estivill X, Ferguson-Smith A, Fitzgibbon J, Flicek P, Giehl C, Graf T, Grosveld F, Guigo R, Gut I, Helin K, Jarvius J, Kuppers R, Lehrach H, Lengauer T, Lernmark A, Leslie D, Loeffler M, Macintyre E, Mai A, Martens JH, Minucci S, Ouwehand WH, Pelicci PG, Pendeville H, Porse B, Rakyan V, Reik W, Schrappe M, Schubeler D, Seifert M, Siebert R, Simmons D, Soranzo N, Spicuglia S, Stratton M, Stunnenberg HG, Tanay A, Torrents D, Valencia A, Vellenga E, Vingron M, Walter J and Willcocks S (2012). BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol.* 30(3):224-6.

Akirav EM, Lebastchi J, Galvan EM, Henegariu O, Akirav M, Ablamunits V, Lizardi PM and Herold KC (2011). Detection of β cell death in diabetes using differentially methylated circulating DNA. *Proc Natl Acad Sci U S A.* 108(47):19018-23.

Allman D and Pillai S (2008). Peripheral B cell subsets. *Curr Opin Immunol.* 20(2):149-57.

Alves JG, Figueiroa JN, Meneses J and Alves GV (2012). Breastfeeding protects against type 1 diabetes mellitus: a case-sibling study. *Breastfeed Med.* 7(1):25-8.

Aly TA, Ide A, Jahromi MM, Barker JM, Fernando MS, Babu SR, Yu L, Miao D, Erlich HA, Fain PR, Barriga KJ, Norris JM, Rewers MJ and Eisenbarth GS (2006). Extreme genetic risk for type 1A diabetes. *Proc Natl Acad Sci U S A.* 103(38):14074-9.

American Diabetes A (2004). Diagnosis and classification of diabetes mellitus. *Diabetes Care.* 27 Suppl 1(S5-S10.

American Diabetes A (2009). Diagnosis and classification of diabetes mellitus. *Diabetes Care.* 32 Suppl 1(S62-7.

Anker P, Lyautey J, Lederrey C and Stroun M (2001). Circulating nucleic acids in plasma or serum. *Clin Chim Acta.* 313(1-2):143-6.

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD and Irizarry RA (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 30(10):1363-9.

Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T and Bock C (2014). Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods.* 11(11):1138-40.

Atkinson MA, Eisenbarth GS and Michels AW (2014). Type 1 diabetes. *Lancet.* 383(9911):69-82.

Augui S, Nora EP and Heard E (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet.* 12(6):429-42.

Bacchetta R, Passerini L, Gambineri E, Dai M, Allan SE, Perroni L, Dagna-Bricarelli F, Sartirana C, Matthes-Martin S, Lawitschka A, Azzari C, Ziegler SF, Levings MK and Roncarolo MG (2006). Defective regulatory and effector T cell functions in patients with FOXP3 mutations. *J Clin Invest.* 116(6):1713-22.

Bach JF (2002). The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med.* 347(12):911-20.

Bach JF (2005). Infections and autoimmune diseases. *J Autoimmun.* 25 Suppl(74-80.

Bach JF and Chatenoud L (2012). The hygiene hypothesis: an explanation for the increased frequency of insulin-dependent diabetes. *Cold Spring Harb Perspect Med.* 2(2):a007799.

Bailey R, Cooper JD, Zeitels L, Smyth DJ, Yang JH, Walker NM, Hyppönen E, Dunger DB, Ramos-Lopez E, Badenhoop K, Nejentsev S and Todd JA (2007). Association of the vitamin D metabolism gene CYP27B1 with type 1 diabetes. *Diabetes.* 56(10):2616-21.

Baranzini SE (2009). The genetics of autoimmune diseases: a networked perspective. *Curr Opin Immunol.* 21(6):596-605.

Barker A, Lauria A, Schloot N, Hosszufalusi N, Ludvigsson J, Mathieu C, Mauricio D, Nordwall M, Van der Schueren B, Mandrup-Poulsen T, Scherbaum WA, Weets I, Gorus FK, Wareham N, Leslie RD and Pozzilli P (2014). Age-dependent decline of beta-cell function in type 1 diabetes after diagnosis: a multi-centre longitudinal study. *Diabetes Obes Metab.* 16(3):262-7.

Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker NM, Rich SS and Type 1 Diabetes Genetics C (2009). Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet.* 41(6):703-7.

Beck R and Lam-Po-Tang PR (1994). Comparison of cord blood and adult blood lymphocyte normal ranges: a possible explanation for decreased severity of graft versus host disease after cord blood transplantation. *Immunol Cell Biol.* 72(5):440-4.

Beck S and Rakyan VK (2008). The methylome: approaches for global DNA methylation profiling. *Trends Genet.* 24(5):231-7.

Bell CG, Teschendorff AE, Rakyan VK, Maxwell AP, Beck S and Savage DA (2010). Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC Med Genomics.* 3(33.

Bell JT and Saffery R (2012). The value of twins in epigenetic epidemiology. *Int J Epidemiol.* 41(1):140-50.

Bell JT and Spector TD (2011). A twin approach to unraveling epigenetics. *Trends Genet.* 27(3):116-25.

Belot MP, Fradin D, Mai N, Le Fur S, Zelenika D, Kerr-Conte J, Pattou F, Lucas B and Bougneres P (2013). CpG methylation changes within the IL2RA promoter in type 1 diabetes of childhood onset. *PLoS One.* 8(7):e68093.

Bennett ST, Lucassen AM, Gough SC, Powell EE, Undlien DE, Pritchard LE, Merriman ME, Kawaguchi Y, Dronsfield MJ and Pociot F (1995). Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat Genet.* 9(3):284-92.

Bennett ST and Todd JA (1996). Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Annu Rev Genet.* 30(343-70.

Bernstein BE, Meissner A and Lander ES (2007). The mammalian epigenome. *Cell.* 128(4):669-81.

Besingi W and Johansson A (2014). Smoke-related DNA methylation changes in the etiology of human disease. *Hum Mol Genet.* 23(9):2290-7.

Beyan H, Buckley LR, Yousaf N, Londei M and Leslie RD (2003). A role for innate immunity in type 1 diabetes? *Diabetes Metab Res Rev.* 19(2):89-100.

Beyan H, Down TA, Ramagopalan SV, Uvebrant K, Nilsson A, Holland ML, Gemma C, Giovannoni G, Boehm BO, Ebers GC, Lernmark A, Cilio CM, Leslie RD and Rakyan VK (2012a). Guthrie card methylomics identifies temporally stable epialleles that are present at birth in humans. *Genome Res.* 22(11):2138-45.

Beyan H, Drexhage RC, van der Heul Nieuwenhuijsen L, de Wit H, Padmos RC, Schloot NC, Drexhage HA and Leslie RD (2010). Monocyte gene-expression profiles associated with childhood-onset type 1 diabetes and disease risk: a study of identical twins. *Diabetes.* 59(7):1751-5.

Beyan H, Riese H, Hawa MI, Beretta G, Davidson HW, Hutton JC, Burger H, Schlosser M, Snieder H, Boehm BO and Leslie RD (2012b). Glycotoxin and autoantibodies are additive environmentally determined predictors of type 1 diabetes: a twin and population study. *Diabetes.* 61(5):1192-8.

Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R and Gunderson KL (2009). Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics.* 1(1):177-200.

Biel M, Wascholowski V and Giannis A (2005). Epigenetics--an epicenter of gene regulation: histones and histone-modifying enzymes. *Angew Chem Int Ed Engl.* 44(21):3186-216.

Bingley PJ, Douek IF, Rogers CA and Gale EA (2000). Influence of maternal age at delivery and birth order on risk of type 1 diabetes in childhood: prospective population based family study. Bart's-Oxford Family Study Group. *BMJ.* 321(7258):420-4.

Bird A (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16(1):6-21.

Bird A (2007). Perceptions of epigenetics. *Nature.* 447(7143):396-8.

Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstrom TJ, Harris TB, Launer LJ, Eiriksdottir G, Leppert MF, Sapienza C, Gudnason V and Feinberg AP (2008). Intra-individual change over time in DNA methylation with familial clustering. *JAMA.* 299(24):2877-83.

Blattler A, Yao L, Witt H, Guo Y, Nicolet CM, Berman BP and Farnham PJ (2014). Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol.* 15(9):469.

Bluestone JA, Herold K and Eisenbarth G (2010). Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature.* 464(7293):1293-300.

Bock C (2012). Analysing and interpreting DNA methylation data. *Nat Rev Genet.* 13(10):705-19.

Bock C, Paulsen M, Tierling S, Mikeska T, Lengauer T and Walter J (2006). CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats, and predicted DNA structure. *PLoS Genet.* 2(3):e26.

Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG and Meissner A (2010). Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol.* 28(10):1106-14.

Bodansky HJ, Staines A, Stephenson C, Haigh D and Cartwright R (1992). Evidence for an environmental effect in the aetiology of insulin dependent diabetes in a transmigratory population. *BMJ.* 304(6833):1020-2.

Boomsma D, Busjahn A and Peltonen L (2002). Classical twin studies and beyond. *Nat Rev Genet.* 3(11):872-82.

Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M, Eisenbarth GS, Comings D and Mustelin T (2004). A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet.* 36(4):337-8.

Bradshaw EM, Raddassi K, Elyaman W, Orban T, Gottlieb PA, Kent SC and Hafler DA (2009). Monocytes from patients with type 1 diabetes spontaneously secrete proinflammatory cytokines inducing Th17 cells. *J Immunol.* 183(7):4432-9.

Bramswig NC and Kaestner KH (2012). Epigenetics and diabetes treatment: an unrealized promise? *Trends Endocrinol Metab.* 23(6):286-91.

Branco MR, Ficz G and Reik W (2011). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet.*

Brasacchio D, Okabe J, Tikellis C, Balcerczyk A, George P, Baker EK, Calkin AC, Brownlee M, Cooper ME and El-Osta A (2009). Hyperglycemia induces a dynamic cooperativity of histone methylase and demethylase enzymes associated with gene-activating epigenetic marks that coexist on the lysine tail. *Diabetes.* 58(5):1229-36.

Breitling LP, Yang R, Korn B, Burwinkel B and Brenner H (2011). Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet.* 88(4):450-7.

Brennan EP, Ehrich M, Brazil DP, Crean JK, Murphy M, Sadlier DM, Martin F, Godson C, McKnight AJ, van den Boom D, Maxwell AP and Savage DA (2009). Comparative analysis of DNA methylation profiles in peripheral blood leukocytes versus lymphoblastoid cell lines. *Epigenetics.* 4(3):159-64.

Brooks WH, Le Dantec C, Pers JO, Youinou P and Renaudineau Y (2010). Epigenetics and autoimmunity. *J Autoimmun.* 34(3):J207-19.

Brophy S, Yderstraede K, Mauricio D, Hunter S, Hawa M, Pozzilli P, Schernthaner G, Schloot N, Buzzetti R, Davies H, Leslie D, Williams R and Action LG (2008). Time to insulin initiation cannot be used in defining latent autoimmune diabetes in adults. *Diabetes Care.* 31(3):439-41.

Brusko TM, Wasserfall CH, Clare-Salzler MJ, Schatz DA and Atkinson MA (2005). Functional defects and the influence of age on the frequency of CD4+ CD25+ T-cells in type 1 diabetes. *Diabetes.* 54(5):1407-14.

Caramori ML, Kim Y, Moore JH, Rich SS, Mychaleckyj JC, Kikyo N and Mauer M (2012). Gene expression differences in skin fibroblasts in identical twins discordant for type 1 diabetes. *Diabetes.* 61(3):739-44.

Carone BR, Fauquier L, Habib N, Shea JM, Hart CE, Li R, Bock C, Li C, Gu H, Zamore PD, Meissner A, Weng Z, Hofmann HA, Friedman N and Rando OJ (2010). Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell.* 143(7):1084-96.

Chen Z, Fei M, Fu D, Zhang L, Ma Y, Wang Y, Zhang F, Xia Q and Wang X (2013). Association between cytotoxic T lymphocyte antigen-4 polymorphism and type 1 diabetes: a meta-analysis. *Gene.* 516(2):263-70.

Christensen DP, Gysemans C, Lundh M, Dahllof MS, Noesgaard D, Schmidt SF, Mandrup S, Birkbak N, Workman CT, Piemonti L, Blaabjerg L, Monzani V, Fossati G, Mascagni P, Paraskevas S, Aikin RA, Billestrup N, Grunnet LG, Dinarello CA, Mathieu C and Mandrup-Poulsen T (2014). Lysine deacetylase inhibition prevents diabetes by chromatin-independent immunoregulation and beta-cell protection. *Proc Natl Acad Sci U S A.* 111(3):1055-9.

Cinek O, Stene LC, Kramna L, Tapia G, Oikarinen S, Witso E, Rasmussen T, Torjesen PA, Hyoty H and Ronningen KS (2014). Enterovirus RNA in longitudinal blood samples and risk of islet autoimmunity in children with a high genetic risk of type 1 diabetes: the MIDIA study. *Diabetologia.* 57(10):2193-200.

Clements GB, Galbraith DN and Taylor KW (1995). Coxsackie B virus infection and onset of childhood diabetes. *Lancet.* 346(8969):221-3.

Cnop M, Welsh N, Jonas JC, Jörns A, Lenzen S and Eizirik DL (2005). Mechanisms of pancreatic beta-cell death in type 1 and type 2 diabetes: many differences, few similarities. *Diabetes.* 54 Suppl 2(S97-107.

Consortium EP (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature.* 489(7414):57-74.

Cooney CA, Dave AA and Wolff GL (2002). Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J Nutr.* 132(8 Suppl):2393S-2400S.

Cooper JD, Smyth DJ, Walker NM, Stevens H, Burren OS, Wallace C, Greissl C, Ramos-Lopez E, Hyppönen E, Dunger DB, Spector TD, Ouwehand WH, Wang TJ, Badenhoop K and Todd JA (2011). Inherited variation in vitamin D genes is associated with predisposition to autoimmune disease type 1 diabetes. *Diabetes.* 60(5):1624-31.

Correa-Giannella ML and Raposo do Amaral AS (2009). Pancreatic islet transplantation. *Diabetol Metab Syndr.* 1(1):9.

Craig ME, Nair S, Stein H and Rawlinson WD (2013). Viruses and type 1 diabetes: a new look at an old story. *Pediatr Diabetes.* 14(3):149-58.

D'Cruz D (2000). Autoimmune diseases associated with drugs, chemicals and environmental factors. *Toxicol Lett.* 112-113(421-32.

Dang M, Rockell J, Wagner R, Wenzlau JM, Yu L, Hutton JC, Gottlieb PA and Davidson HW (2011). Human type 1 diabetes is associated with T cell autoimmunity to zinc transporter 8. *J Immunol.* 186(10):6056-63.

Dang MN, Buzzetti R and Pozzilli P (2013). Epigenetics in autoimmune diseases with focus on type 1 diabetes. *Diabetes Metab Res Rev.* 29(1):8-18.

Davis AK, DuBose SN, Haller MJ, Miller KM, DiMeglio LA, Bethin KE, Goland RS, Greenberg EM, Liljenquist DR, Ahmann AJ, Marcovina SM, Peters AL, Beck RW, Greenbaum CJ and for the TDECN (2014). Prevalence of Detectable C-peptide According to Age at Diagnosis and Duration of Type 1 Diabetes. *Diabetes Care.*

de Almeida MC, Silva AC, Barral A and Barral Netto M (2000). A simple method for human peripheral blood monocyte isolation. *Mem Inst Oswaldo Cruz.* 95(2):221-3.

De Filippo G, Pozzi N, Cosentini E, Cavalcanti M, Carel JC, Tamasi S, Franzese A and Pignata C (1997). Increased CD5+CD19+ B lymphocytes at the onset of type 1 diabetes in children. *Acta Diabetol.* 34(4):271-4.

de Laat W and Duboule D (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature.* 502(7472):499-506.

Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C and Fuks F (2011). Evaluation of the Infinium Methylation 450K technology. *Epigenomics.* 3(6):771-84.

Depoil D, Fleire S, Treanor BL, Weber M, Harwood NE, Marchbank KL, Tybulewicz VL and Batista FD (2008). CD19 is essential for B cell activation by promoting B cell receptor-antigen microcluster formation in response to membrane-bound ligand. *Nat Immunol.* 9(1):63-72.

Dotta F, Censini S, van Halteren AG, Marselli L, Masini M, Dionisi S, Mosca F, Boggi U, Muda AO, Del Prato S, Elliott JF, Covacci A, Rappuoli R, Roep BO and Marchetti P (2007). Coxsackie B4 virus infection of beta cells and natural killer cell insulitis in recent-onset type 1 diabetic patients. *Proc Natl Acad Sci U S A.* 104(12):5115-20.

Downes K, Pekalski M, Angus KL, Hardy M, Nutland S, Smyth DJ, Walker NM, Wallace C and Todd JA (2010). Reduced expression of IFIH1 is protective for type 1 diabetes. *PLoS One.* 5(9).

Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L and Lin SM (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics.* 11(587.

Duckworth WC, Bennett RG and Hamel FG (1998). Insulin degradation: progress and potential. *Endocr Rev.* 19(5):608-24.

Dunne JL, Triplett EW, Gevers D, Xavier R, Insel R, Danska J and Atkinson MA (2014). The intestinal microbiome in type 1 diabetes. *Clin Exp Immunol.* 177(1):30-7.

Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, Danenberg PV and Laird PW (2000). MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* 28(8):E32.

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K and Beck S (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet.* 38(12):1378-85.

Eizirik DL and Mandrup-Poulsen T (2001). A choice of death--the signal-transduction of immune-mediated beta-cell apoptosis. *Diabetologia.* 44(12):2115-33.

El Gazzar M and McCall CE (2011). MicroRNAs regulatory networks in myeloid lineage development and differentiation: regulators of the regulators. *Immunol Cell Biol.*

El Hajj N, Pliushch G, Schneider E, Dittrich M, Muller T, Korenkov M, Aretz M, Zechner U, Lehnen H and Haaf T (2013). Metabolic programming of MEST DNA methylation by intrauterine exposure to gestational diabetes mellitus. *Diabetes.* 62(4):1320-8.

Elahi E and Ronaghi M (2004). Pyrosequencing: a tool for DNA sequencing analysis. *Methods Mol Biol.* 255(211-9.

EURODIAB ASG (2000). Variation and trends in incidence of childhood diabetes in Europe. EURODIAB ACE Study Group. *Lancet.* 355(9207):873-6.

Farber DL, Yudanin NA and Restifo NP (2014). Human memory T cells: generation, compartmentalization and homeostasis. *Nat Rev Immunol.* 14(1):24-35.

Feinberg AP (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature.* 447(7143):433-40.

Fernandez-Valverde SL, Taft RJ and Mattick JS (2011). MicroRNAs in β-cell biology, insulin resistance, diabetes and its complications. *Diabetes.* 60(7):1825-31.

Fettke F, Schumacher A, Costa SD and Zenclussen AC (2014). B cells: the old new players in reproductive immunology. *Front Immunol.* 5(285.

Field LL (2002). Genetic linkage and association studies of Type I diabetes: challenges and rewards. *Diabetologia.* 45(1):21-35.

Fisher MM, Perez Chumbiauca CN, Mather KJ, Mirmira RG and Tersey SA (2013). Detection of islet beta-cell death in vivo by multiplex PCR analysis of differentially methylated DNA. *Endocrinology.* 154(9):3476-81.

Fitzner K and Heckinger E (2010). Sample size calculation and power analysis: a quick review. *Diabetes Educ.* 36(5):701-7.

Floess S, Freyer J, Siewert C, Baron U, Olek S, Polansky J, Schlawe K, Chang HD, Bopp T, Schmitt E, Klein-Hessling S, Serfling E, Hamann A and Huehn J (2007). Epigenetic control of the foxp3 locus in regulatory T cells. *PLoS Biol.* 5(2):e38.

Fradin D, Le Fur S, Mille C, Naoui N, Groves C, Zelenika D, McCarthy MI, Lathrop M and Bougneres P (2012). Association of the CpG methylation pattern of the proximal insulin gene promoter with type 1 diabetes. *PLoS One.* 7(5):e36278.

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C and Esteller M (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A.* 102(30):10604-9.

Fraga MF and Esteller M (2007). Epigenetics and aging: the targets and the marks. *Trends Genet.* 23(8):413-8.

Gardiner-Garden M and Frommer M (1987). CpG islands in vertebrate genomes. *J Mol Biol.* 196(2):261-82.

Garg G, Tyler JR, Yang JH, Cutler AJ, Downes K, Pekalski M, Bell GL, Nutland S, Peakman M, Todd JA, Wicker LS and Tree TI (2012). Type 1 diabetes-associated IL2RA variation lowers IL-2 signaling and contributes to diminished CD4+CD25+ regulatory T cell function. *J Immunol.* 188(9):4644-53.

Garibyan L and Avashia N (2013). Polymerase chain reaction. *J Invest Dermatol.* 133(3):e6.

Geissmann F, Auffray C, Palframan R, Wirrig C, Ciocca A, Campisi L, Narni-Mancinelli E and Lauvau G (2008). Blood monocytes: distinct subsets, how they relate to dendritic cells, and their possible roles in the regulation of T-cell responses. *Immunol Cell Biol.* 86(5):398-408.

Geissmann F, Manz MG, Jung S, Sieweke MH, Merad M and Ley K (2010). Development of monocytes, macrophages, and dendritic cells. *Science.* 327(5966):656-61.

Germain RN (2002). T-cell development and the CD4-CD8 lineage decision. *Nat Rev Immunol.* 2(5):309-22.

Gillespie KM, Bain SC, Barnett AH, Bingley PJ, Christie MR, Gill GV and Gale EA (2004). The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. *Lancet.* 364(9446):1699-700.

Giongo A, Gano KA, Crabb DB, Mukherjee N, Novelo LL, Casella G, Drew JC, Ilonen J, Knip M, Hyoty H, Veijola R, Simell T, Simell O, Neu J, Wasserfall CH,

Schatz D, Atkinson MA and Triplett EW (2011). Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J.* 5(1):82-91.

Gonzalez-Mejia ME and Doseff AI (2009). Regulation of monocytes and macrophages cell fate. *Front Biosci (Landmark Ed).* 14(2413-31.

Goodnow CC, Sprent J, Fazekas de St Groth B and Vinuesa CG (2005). Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature.* 435(7042):590-7.

Gordon S and Taylor PR (2005). Monocyte and macrophage heterogeneity. *Nat Rev Immunol.* 5(12):953-64.

Gout J, Pommier RM, Vincent DF, Ripoche D, Goddard-Leon S, Colombe A, Treilleux I, Valcourt U, Tomasini R, Dufresne M, Bertolino P and Bartholin L (2013). The conditional expression of KRAS G12D in mouse pancreas induces disorganization of endocrine islets prior the onset of ductal pre-cancerous lesions. *Pancreatology.* 13(3):191-5.

Graham J, Kockum I, Sanjeevi CB, Landin-Olsson M, Nystrom L, Sundkvist G, Arnqvist H, Blohme G, Lithner F, Littorin B, Schersten B, Wibell L, Ostman J, Lernmark A, Breslow N and Dahlquist G (1999). Negative association between type 1 diabetes and HLA DQB1*0602-DQA1*0102 is attenuated with age at onset. Swedish Childhood Diabetes Study Group. *Eur J Immunogenet.* 26(2-3):117-27.

Gregersen PK and Behrens TW (2006). Genetics of autoimmune diseases--disorders of immune homeostasis. *Nat Rev Genet.* 7(12):917-28.

Gregg BE, Moore PC, Demozay D, Hall BA, Li M, Husain A, Wright AJ, Atkinson MA and Rhodes CJ (2012). Formation of a human beta-cell population within pancreatic islets is set early in life. *J Clin Endocrinol Metab.* 97(9):3197-206.

Grieco FA, Vendrame F, Spagnuolo I and Dotta F (2011). Innate immunity and the pathogenesis of type 1 diabetes. *Semin Immunopathol.* 33(1):57-66.

Guay C, Roggli E, Nesca V, Jacovetti C and Regazzi R (2011). Diabetes mellitus, a microRNA-related disease? *Transl Res.* 157(4):253-64.

Ha TY (2011). The Role of MicroRNAs in Regulatory T Cells and in the Immune Response. *Immune Netw.* 11(1):11-41.

Hagen KB, Byfuglien MG, Falzon L, Olsen SU and Smedslund G (2009). Dietary interventions for rheumatoid arthritis. *Cochrane Database Syst Rev.* 1):CD006400.

Han D, Cai X, Wen J, Matheson D, Skyler JS, Kenyon NS and Chen Z (2012). Innate and adaptive immune gene expression profiles as biomarkers in human type 1 diabetes. *Clin Exp Immunol.* 170(2):131-8.

Han L and Zhao Z (2009). CpG islands or CpG clusters: how to identify functional GC-rich regions in a genome? *BMC Bioinformatics.* 10(65.

Hansen KD, Langmead B and Irizarry RA (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* 13(10):R83.

Hansen KD, Sabunciyan S, Langmead B, Nagy N, Curley R, Klein G, Klein E, Salamon D and Feinberg AP (2014). Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome Res.* 24(2):177-84.

Hansen TF, Alvarez-Castro JM, Carter AJ, Hermisson J and Wagner GP (2006). Evolution of genetic architecture under directional selection. *Evolution.* 60(8):1523-36.

Hardison RC and Taylor J (2012). Genomic approaches towards finding cis-regulatory modules in animals. *Nat Rev Genet.* 13(7):469-83.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, Olshen A, Ballinger T, Zhou X, Forsberg KJ, Gu J, Echipare L, O'Geen H, Lister R, Pelizzola M, Xi Y, Epstein CB, Bernstein BE, Hawkins RD, Ren B, Chung WY, Gu H, Bock C, Gnirke A, Zhang MQ, Haussler D, Ecker JR, Li W, Farnham PJ, Waterland RA, Meissner A, Marra MA, Hirst M, Milosavljevic A and Costello JF (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol.* 28(10):1097-105.

Harron KL, Feltbower RG, McKinney PA, Bodansky HJ, Campbell FM and Parslow RC (2011a). Rising rates of all types of diabetes in south Asian and non-south Asian children and young people aged 0-29 years in West Yorkshire, U.K., 1991-2006. *Diabetes Care.* 34(3):652-4.

Harron KL, McKinney PA, Feltbower RG, Bodansky HJ, Norman PD, Campbell FM and Parslow RC (2011b). Incidence rate trends in childhood type 1 diabetes in Yorkshire, UK 1978-2007: effects of deprivation and age at diagnosis in the South Asian and non-South Asian populations. *Diabet Med.* 28(12):1508-13.

Hart JE, Laden F, Puett RC, Costenbader KH and Karlson EW (2009). Exposure to traffic pollution and increased risk of rheumatoid arthritis. *Environ Health Perspect.* 117(7):1065-9.

Hawa M, Rowe R, Lan MS, Notkins AL, Pozzilli P, Christie MR and Leslie RD (1997). Value of antibodies to islet protein tyrosine phosphatase-like molecule in predicting type 1 diabetes. *Diabetes.* 46(8):1270-5.

Hawa MI, Kolb H, Schloot N, Beyan H, Paschou SA, Buzzetti R, Mauricio D, De Leiva A, Yderstraede K, Beck-Neilsen H, Tuomilehto J, Sarti C, Thivolet C, Hadden D, Hunter S, Schernthaner G, Scherbaum WA, Williams R, Brophy S, Pozzilli P, Leslie RD and Action Lc (2013). Adult-onset autoimmune diabetes in Europe is prevalent with a broad clinical phenotype: Action LADA 7. *Diabetes Care.* 36(4):908-13.

Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE and Lumey LH (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A.* 105(44):17046-9.

Heikkinen SM, Pitkaniemi JM, Kilpelainen ML and Koskenvuo MJ (2013). Does farm environment protect against type 1 diabetes mellitus? *Diab Vasc Dis Res.* 10(4):375-7.

Heinig M, Petretto E, Wallace C, Bottolo L, Rotival M, Lu H, Li Y, Sarwar R, Langley SR, Bauerfeind A, Hummel O, Lee YA, Paskas S, Rintisch C, Saar K, Cooper J, Buchan R, Gray EE, Cyster JG, Erdmann J, Hengstenberg C, Maouche S, Ouwehand WH, Rice CM, Samani NJ, Schunkert H, Goodall AH, Schulz H, Roider HG, Vingron M, Blankenberg S, Münzel T, Zeller T, Szymczak S, Ziegler A, Tiret L, Smyth DJ, Pravenec M, Aitman TJ, Cambien F, Clayton D, Todd JA, Hubner N, Cook SA and Consortium C (2010). A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature.* 467(7314):460-4.

Heinz S, Romanoski CE, Benner C, Allison KA, Kaikkonen MU, Orozco LD and Glass CK (2013). Effect of natural genetic variation on enhancer selection and function. *Nature.* 503(7477):487-92.

Herman JG, Graff JR, Myohanen S, Nelkin BD and Baylin SB (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A.* 93(18):9821-6.

Hermann R, Knip M, Veijola R, Simell O, Laine AP, Akerblom HK, Groop PH, Forsblom C, Pettersson-Fernholm K, Ilonen J and FinnDiane Study G (2003). Temporal changes in the frequencies of HLA genotypes in patients with Type 1 diabetes--indication of an increased environmental pressure? *Diabetologia.* 46(3):420-5.

Herold KC, Vignali DA, Cooke A and Bluestone JA (2013). Type 1 diabetes: translating mechanistic observations into effective clinical outcomes. *Nat Rev Immunol.* 13(4):243-56.

Herzenberg LA, Tung J, Moore WA, Herzenberg LA and Parks DR (2006). Interpreting flow cytometry data: a guide for the perplexed. *Nat Immunol.* 7(7):681-5.

Hewagama A and Richardson B (2009). The genetics and epigenetics of autoimmune diseases. *J Autoimmun.* 33(1):3-11.

Heyn H and Esteller M (2012). DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet.* 13(10):679-92.

Hezova R, Slaby O, Faltejskova P, Mikulkova Z, Buresova I, Raja KR, Hodek J, Ovesna J and Michalek J (2010). microRNA-342, microRNA-191 and microRNA-510 are differentially expressed in T regulatory cells of type 1 diabetic patients. *Cell Immunol.* 260(2):70-4.

Hinman RM and Cambier JC (2014). Role of B lymphocytes in the pathogenesis of type 1 diabetes. *Curr Diab Rep.* 14(11):543.

Hitman GA, Mannan N, McDermott MF, Aganna E, Ogunkolade BW, Hales CN and Boucher BJ (1998). Vitamin D receptor gene polymorphisms influence insulin secretion in Bangladeshi Asians. *Diabetes.* 47(4):688-90.

Holdenrieder S, Nagel D, Schalhorn A, Heinemann V, Wilkowski R, von Pawel J, Raith H, Feldmann K, Kremer AE, Muller S, Geiger S, Hamann GF, Seidel D and Stieber P (2008). Clinical relevance of circulating nucleosomes in cancer. *Ann N Y Acad Sci.* 1137(180-9.

Holdenrieder S, Stieber P, Chan LY, Geiger S, Kremer A, Nagel D and Lo YM (2005). Cell-free DNA in serum and plasma: comparison of ELISA and quantitative PCR. *Clin Chem.* 51(8):1544-6.

Holland PM, Abramson RD, Watson R and Gelfand DH (1991). Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A.* 88(16):7276-80.

Hon GC, Hawkins RD and Ren B (2009). Predictive chromatin signatures in the mammalian genome. *Hum Mol Genet.* 18(R2):R195-201.

Horvath S (2013). DNA methylation age of human tissues and cell types. *Genome Biol.* 14(10):R115.

Houseman EA, Molitor J and Marsit CJ (2014). Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics.* 30(10):1431-9.

Howson JM, Rosinger S, Smyth DJ, Boehm BO, Group A-ES and Todd JA (2011). Genetic analysis of adult-onset autoimmune diabetes. *Diabetes.* 60(10):2645-53.

Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC and Lempicki RA (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* 35(Web Server issue):W169-75.

Huang J, Ellinghaus D, Franke A, Howie B and Li Y (2012). 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *Eur J Hum Genet.* 20(7):801-5.

Huber A, Menconi F, Corathers S, Jacobson EM and Tomer Y (2008). Joint genetic susceptibility to type 1 diabetes and autoimmune thyroiditis: from epidemiology to mechanisms. *Endocr Rev.* 29(6):697-725.

Hummel S, Vehik K, Uusitalo U, McLeod W, Aronsson CA, Frank N, Gesualdo P, Yang J, Norris JM, Virtanen SM and Group TS (2014). Infant feeding patterns in families with a diabetes history - observations from The Environmental Determinants of Diabetes in the Young (TEDDY) birth cohort study. *Public Health Nutr.* 17(12):2853-62.

Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, Barrett JC, Blackburn H, Brand O, Burren O, Capon F, Compston A, Gough SC, Jostins L, Kong Y, Lee JC, Lek M, MacArthur DG, Mansfield JC, Mathew CG, Mein CA, Mirza M, Nutland S, Onengut-Gumuscu S, Papouli E, Parkes M, Rich SS, Sawcer S, Satsangi J, Simmonds MJ, Trembath RC, Walker NM, Wozniak E, Todd JA, Simpson MA, Plagnol V and van Heel DA (2013). Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature.* 498(7453):232-5.

Husain Z, Kelly MA, Eisenbarth GS, Pugliese A, Awdeh ZL, Larsen CE and Alper CA (2008). The MHC type 1 diabetes susceptibility gene is centromeric to HLA-DQB1. *J Autoimmun.* 30(4):266-72.

Husseiny MI, Kaye A, Zebadua E, Kandeel F and Ferreri K (2014). Tissue-specific methylation of human insulin gene and PCR assay for monitoring Beta cell death. *PLoS One.* 9(4):e94591.

Husseiny MI, Kuroda A, Kaye AN, Nair I, Kandeel F and Ferreri K (2012). Development of a quantitative methylation-specific polymerase chain reaction method for monitoring beta cell death in type 1 diabetes. *PLoS One.* 7(10):e47942.

Hussen HI, Persson M and Moradi T (2013). The trends and the risk of type 1 diabetes over the past 40 years: an analysis by birth cohorts and by parental migration background in Sweden. *BMJ Open.* 3(10):e003418.

Hyöty H and Taylor KW (2002). The role of viruses in human diabetes. *Diabetologia.* 45(10):1353-61.

Hyppönen E, Läärä E, Reunanen A, Järvelin MR and Virtanen SM (2001). Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. *Lancet.* 358(9292):1500-3.

Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M and Tuomilehto J (2003). Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes.* 52(4):1052-5.

Ichii M, Oritani K and Kanakura Y (2014). Early B lymphocyte development: Similarities and differences in human and mouse. *World J Stem Cells.* 6(4):421-31.

Ikegami H, Fujisawa T, Kawabata Y, Noso S and Ogihara T (2006). Genetics of type 1 diabetes: similarities and differences between Asian and Caucasian populations. *Ann N Y Acad Sci.* 1079(51-9.

Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R and Bird AP (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* 6(9):e1001134.

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Sabunciyan S and Feinberg AP (2009).

The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet.* 41(2):178-86.

Irvine KM, Gallego P, An X, Best SE, Thomas G, Wells C, Harris M, Cotterill A and Thomas R (2012). Peripheral blood monocyte gene expression profile clinically stratifies patients with recent-onset type 1 diabetes. *Diabetes.* 61(5):1281-90.

Itoh N, Hanafusa T, Miyazaki A, Miyagawa J, Yamagata K, Yamamoto K, Waguri M, Imagawa A, Tamura S and Inada M (1993). Mononuclear cell infiltration and its relation to the expression of major histocompatibility complex antigens and adhesion molecules in pancreas biopsy specimens from newly diagnosed insulin-dependent diabetes mellitus patients. *J Clin Invest.* 92(5):2313-22.

Jaenisch R and Bird A (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet.* 33 Suppl(245-54.

Jaffe AE and Irizarry RA (2014). Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 15(2):R31.

Janeway CA, Jr. (2001). How the immune system protects the host from infection. *Microbes Infect.* 3(13):1167-71.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A and Cheng X (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature.* 449(7159):248-51.

Jiang L, Brackeva B, Ling Z, Kramer G, Aerts JM, Schuit F, Keymeulen B, Pipeleers D, Gorus F and Martens GA (2013). Potential of protein phosphatase inhibitor 1 as biomarker of pancreatic beta-cell injury in vitro and in vivo. *Diabetes.* 62(8):2683-8.

Johnson JS and Evans-Molina C (2014). Translational implications of the beta-cell epigenome in diabetes mellitus. *Transl Res.*

Jones SR, Carley S and Harrison M (2003). An introduction to power and sample size estimation. *Emerg Med J.* 20(5):453-8.

Joo JE, Wong EM, Baglietto L, Jung CH, Tsimiklis H, Park DJ, Wong NC, English DR, Hopper JL, Severi G, Giles GG and Southey MC (2013). The use of DNA from archival dried blood spots with the Infinium HumanMethylation450 array. *BMC Biotechnol.* 13(23.

Jorundsson E, Lumsden JH and Jacobs RM (1999). Rapid staining techniques in cytopathology: a review and comparison of modified protocols for hematoxylin and eosin, Papanicolaou and Romanowsky stains. *Vet Clin Pathol.* 28(3):100-108.

Julier C, Hyer RN, Davies J, Merlin F, Soularue P, Briant L, Cathelineau G, Deschamps I, Rotter JI and Froguel P (1991). Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. *Nature.* 354(6349):155-9.

Kadziela K, Kowalska H, Rymkiewicz-Kluczynska B, Kowalska M, Miszkurka G, Rybczynska J, Wasik M and Pankowska E (2003). Changes in lymphocyte subsets in children with newly diagnosed type 1 diabetes mellitus. *J Pediatr Endocrinol Metab.* 16(2):185-91.

Kahn HS, Morgan TM, Case LD, Dabelea D, Mayer-Davis EJ, Lawrence JM, Marcovina SM, Imperatore G and Group SfDiYS (2009). Association of type 1 diabetes with month of birth among U.S. youth: The SEARCH for Diabetes in Youth Study. *Diabetes Care.* 32(11):2010-5.

Kaizer EC, Glaser CL, Chaussabel D, Banchereau J, Pascual V and White PC (2007). Gene expression in peripheral blood mononuclear cells from children with diabetes. *J Clin Endocrinol Metab.* 92(9):3705-11.

Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GH, Wong AH, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, McRae AF, Visscher PM, Montgomery GW, Gottesman II, Martin NG and Petronis A (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet.* 41(2):240-5.

Kanno Y, Vahedi G, Hirahara K, Singleton K and O'Shea JJ (2012). Transcriptional and epigenetic control of T helper cell specification: molecular mechanisms underlying commitment and plasticity. *Annu Rev Immunol.* 30(707-31.

Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R and Tuomilehto J (2000). Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. *Diabetes Care.* 23(10):1516-26.

Kawabata Y, Ikegami H, Kawaguchi Y, Fujisawa T, Shintani M, Ono M, Nishino M, Uchigata Y, Lee I and Ogihara T (2002). Asian-specific HLA haplotypes reveal heterogeneity of the contribution of HLA-DR and -DQ haplotypes to susceptibility to type 1 diabetes. *Diabetes.* 51(2):545-51.

Ke X, Cortina-Borja M, Silva BC, Lowe R, Rakyan V and Balding D (2013). Integrated analysis of genome-wide genetic and epigenetic association data for identification of disease mechanisms. *Epigenetics.* 8(11):1236-44.

Keating ST and El-Osta A (2012). Chromatin modifications associated with diabetes. *J Cardiovasc Transl Res.* 5(4):399-412.

Keenan HA, Sun JK, Levine J, Doria A, Aiello LP, Eisenbarth G, Bonner-Weir S and King GL (2010). Residual insulin production and pancreatic ss-cell turnover after 50 years of diabetes: Joslin Medalist Study. *Diabetes.* 59(11):2846-53.

Kim DH, Lee JC, Lee MK, Kim KW and Lee MS (2012). Treatment of autoimmune diabetes in NOD mice by Toll-like receptor 2 tolerance in conjunction with dipeptidyl peptidase 4 inhibition. *Diabetologia.* 55(12):3308-17.

Kimpimaki T, Erkkola M, Korhonen S, Kupila A, Virtanen SM, Ilonen J, Simell O and Knip M (2001a). Short-term exclusive breastfeeding predisposes young children with

increased genetic risk of Type I diabetes to progressive beta-cell autoimmunity. *Diabetologia.* 44(1):63-9.

Kimpimaki T, Kupila A, Hamalainen AM, Kukko M, Kulmala P, Savola K, Simell T, Keskinen P, Ilonen J, Simell O and Knip M (2001b). The first signs of beta-cell autoimmunity appear in infancy in genetically susceptible children from the general population: the Finnish Type 1 Diabetes Prediction and Prevention Study. *J Clin Endocrinol Metab.* 86(10):4782-8.

Klein D, Misawa R, Bravo-Egana V, Vargas N, Rosero S, Piroso J, Ichii H, Umland O, Zhijie J, Tsinoremas N, Ricordi C, Inverardi L, Dominguez-Bendala J and Pastori RL (2013). MicroRNA expression in alpha and beta cells of human pancreatic islets. *PLoS One.* 8(1):e55064.

Klose RJ and Bird AP (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci.* 31(2):89-97.

Knip M, Akerblom HK, Becker D, Dosch HM, Dupre J, Fraser W, Howard N, Ilonen J, Krischer JP, Kordonouri O, Lawson ML, Palmer JP, Savilahti E, Vaarala O, Virtanen SM and Group TS (2014). Hydrolyzed infant formula and early beta-cell autoimmunity: a randomized clinical trial. *JAMA.* 311(22):2279-87.

Knip M, Virtanen SM and Akerblom HK (2010). Infant feeding and the risk of type 1 diabetes. *Am J Clin Nutr.* 91(5):1506S-1513S.

Koch L (2014). Epigenetics: an epigenetic twist on the missing heritability of complex traits. *Nat Rev Genet.* 15(4):218.

Koch MW, Metz LM, Agrawal SM and Yong VW (2013). Environmental factors and their regulation of immunity in multiple sclerosis. *J Neurol Sci.* 324(1-2):10-6.

Koestler DC, Christensen B, Karagas MR, Marsit CJ, Langevin SM, Kelsey KT, Wiencke JK and Houseman EA (2013). Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. *Epigenetics.* 8(8):816-26.

Kormendy D, Hoff H, Hoff P, Broker BM, Burmester GR and Brunner-Weinzierl MC (2013). Impact of the CTLA-4/CD28 axis on the processes of joint inflammation in rheumatoid arthritis. *Arthritis Rheum.* 65(1):81-7.

Kronenberg M and Rudensky A (2005). Regulation of immunity by self-reactive T cells. *Nature.* 435(7042):598-604.

Krueger F, Kreck B, Franke A and Andrews SR (2012). DNA methylome analysis using short bisulfite sequencing data. *Nat Methods.* 9(2):145-51.

Kumar M, Nath S, Prasad HK, Sharma GD and Li Y (2012). MicroRNAs: a new ray of hope for diabetes mellitus. *Protein Cell.* 3(10):726-38.

Kupila A, Keskinen P, Simell T, Erkkila S, Arvilommi P, Korhonen S, Kimpimaki T, Sjoroos M, Ronkainen M, Ilonen J, Knip M and Simell O (2002). Genetic risk determines the emergence of diabetes-associated autoantibodies in young children. *Diabetes.* 51(3):646-51.

Kuroda A, Rauch TA, Todorov I, Ku HT, Al-Abdullah IH, Kandeel F, Mullen Y, Pfeifer GP and Ferreri K (2009). Insulin gene expression is regulated by DNA methylation. *PLoS One.* 4(9):e6953.

Lacobucci G (2013). UK has fifth highest rate of type 1 diabetes in children, new figures show. *BMJ.* 346(f22.

Laird PW (2003). The power and the promise of DNA methylation markers. *Nat Rev Cancer.* 3(4):253-66.

Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE and Kobor MS (2012). Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A.* 109 Suppl 2(17253-60.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ and Snyder M (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22(9):1813-31.

Langevin SM, Houseman EA, Christensen BC, Wiencke JK, Nelson HH, Karagas MR, Marsit CJ and Kelsey KT (2011). The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood. *Epigenetics.* 6(7):908-19.

Larosa DF and Orange JS (2008). 1. Lymphocytes. *J Allergy Clin Immunol.* 121(2 Suppl):S364-9; quiz S412.

Larsson PG, Lakshmikanth T, Svedin E, King C and Flodstrom-Tullberg M (2013). Previous maternal infection protects offspring from enterovirus infection and prevents experimental diabetes development in mice. *Diabetologia.* 56(4):867-74.

Laurberg P, Pedersen KM, Vestergaard H and Sigurdsson G (1991). High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of Graves' disease in the young in a high iodine intake area: comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. *J Intern Med.* 229(5):415-20.

Lawson JM, Tremble J, Dayan C, Beyan H, Leslie RD, Peakman M and Tree TI (2008). Increased resistance to CD4+CD25hi regulatory T cell-mediated suppression in patients with type 1 diabetes. *Clin Exp Immunol.* 154(3):353-9.

Laybutt DR, Weir GC, Kaneto H, Lebet J, Palmiter RD, Sharma A and Bonner-Weir S (2002). Overexpression of c-Myc in beta-cells of transgenic mice causes proliferation and apoptosis, downregulation of insulin gene expression, and diabetes. *Diabetes.* 51(6):1793-804.

Lebastchi J, Deng S, Lebastchi AH, Beshar I, Gitelman S, Willi S, Gottlieb P, Akirav EM, Bluestone JA and Herold KC (2013). Immune therapy and beta-cell death in type 1 diabetes. *Diabetes.* 62(5):1676-80.

LeBien TW and Tedder TF (2008). B lymphocytes: how they develop and function. *Blood.* 112(5):1570-80.

Lee H, Jaffe AE, Feinberg JI, Tryggvadottir R, Brown S, Montano C, Aryee MJ, Irizarry RA, Herbstman J, Witter FR, Goldman LR, Feinberg AP and Fallin MD (2012a). DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. *Int J Epidemiol.* 41(1):188-99.

Lee PP, Fitzpatrick DR, Beard C, Jessup HK, Lehar S, Makar KW, Perez-Melgosa M, Sweetser MT, Schlissel MS, Nguyen S, Cherry SR, Tsai JH, Tucker SM, Weaver WM, Kelso A, Jaenisch R and Wilson CB (2001). A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity.* 15(5):763-74.

Lee ST, Xiao Y, Muench MO, Xiao J, Fomin ME, Wiencke JK, Zheng S, Dou X, de Smith A, Chokkalingam A, Buffler P, Ma X and Wiemels JL (2012b). A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res.* 40(22):11339-51.

Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K and Irizarry RA (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet.* 11(10):733-9.

Leslie D, Lipsky P and Notkins AL (2001). Autoantibodies as predictors of disease. *J Clin Invest.* 108(10):1417-22.

Leslie RD (2010). Predicting adult-onset autoimmune diabetes: clarity from complexity. *Diabetes.* 59(2):330-1.

Leslie RD, Atkinson MA and Notkins AL (1999). Autoantigens IA-2 and GAD in Type I (insulin-dependent) diabetes. *Diabetologia.* 42(1):3-14.

Leslie RD and Delli Castelli M (2004). Age-dependent influences on the origins of autoimmune diabetes: evidence and implications. *Diabetes.* 53(12):3033-40.

Leslie RD, Kolb H, Schloot NC, Buzzetti R, Mauricio D, De Leiva A, Yderstraede K, Sarti C, Thivolet C, Hadden D, Hunter S, Schernthaner G, Scherbaum W, Williams R and Pozzilli P (2008). Diabetes classification: grey zones, sound and smoke: Action LADA 1. *Diabetes Metab Res Rev.* 24(7):511-9.

Lettre G and Rioux JD (2008). Autoimmune diseases: insights from genome-wide association studies. *Hum Mol Genet.* 17(R2):R116-21.

Li E (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet.* 3(9):662-73.

Li LC and Dahiya R (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics.* 18(11):1427-31.

Li Y and Tollefsbol TO (2011). DNA methylation detection: bisulfite genomic sequencing analysis. *Methods Mol Biol.* 791(11-21.

Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, Yu J, Wu H, Sun J, Zhang H, Chen Q, Luo R, Chen M, He Y, Jin X, Zhang Q, Yu C, Zhou G, Sun J, Huang Y, Zheng H, Cao H, Zhou X, Guo S, Hu X, Li X, Kristiansen K, Bolund L, Xu J, Wang W, Yang H, Wang J, Li R, Beck S, Wang J and Zhang X (2010). The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol.* 8(11):e1000533.

Lie BA, Todd JA, Pociot F, Nerup J, Akselsen HE, Joner G, Dahl-Jorgensen K, Ronningen KS, Thorsby E and Undlien DE (1999). The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. *Am J Hum Genet.* 64(3):793-800.

Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, Sigvardsson M, Hagman J, Espinoza CA, Dutkowski J, Ideker T, Glass CK and Murre C (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol.* 11(7):635-43.

Lindley S, Dayan CM, Bishop A, Roep BO, Peakman M and Tree TI (2005). Defective suppressor function in CD4(+)CD25(+) T-cells from patients with type 1 diabetes. *Diabetes.* 54(1):92-9.

Lister R and Ecker JR (2009). Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res.* 19(6):959-66.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L and Law M (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol.* 2012(251364.

Liu L, Li Y and Tollefsbol TO (2008). Gene-environment interactions and epigenetic basis of human diseases. *Curr Issues Mol Biol.* 10(1-2):25-36.

Lönnrot M, Korpela K, Knip M, Ilonen J, Simell O, Korhonen S, Savola K, Muona P, Simell T, Koskela P and Hyöty H (2000). Enterovirus infection as a risk factor for beta-cell autoimmunity in a prospectively observed birth cohort: the Finnish Diabetes Prediction and Prevention Study. *Diabetes.* 49(8):1314-8.

Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, Bourget K, Plagnol V, Field S, Atkinson M, Clayton DG, Wicker LS and Todd JA (2007). Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Genet.* 39(9):1074-82.

Lowe R, Gemma C, Beyan H, Hawa MI, Bazeos A, Leslie RD, Montpetit A, Rakyan VK and Ramagopalan SV (2013). Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics.* 8(4).

Lowe R and Rakyan VK (2013). Marmal-aid--a database for Infinium HumanMethylation450. *BMC Bioinformatics.* 14(359.

Lowe R and Rakyan VK (2014). Correcting for cell-type composition bias in epigenome-wide association studies. *Genome Med.* 6(3):23.

Lu Q (2013). The critical importance of epigenetics in autoimmunity. *J Autoimmun.* 41(1-5.

Lucassen AM, Julier C, Beressi JP, Boitard C, Froguel P, Lathrop M and Bell JI (1993). Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nat Genet.* 4(3):305-10.

Luckheeram RV, Zhou R, Verma AD and Xia B (2012). CD4(+)T cells: differentiation and functions. *Clin Dev Immunol.* 2012(925135.

Lund-Blix NA, Stene LC, Rasmussen T, Torjesen PA, Andersen LF and Ronningen KS (2014). Infant Feeding in Relation to Islet Autoimmunity and Type 1 Diabetes in Genetically Susceptible Children: The MIDIA Study. *Diabetes Care.*

Lyons PA, Koukoulaki M, Hatton A, Doggett K, Woffendin HB, Chaudhry AN and Smith KG (2007). Microarray analysis of human leucocyte subsets: the advantages of positive selection and rapid purification. *BMC Genomics.* 8(64.

Mallone R, Mannering SI, Brooks-Worrell BM, Durinovic-Bello I, Cilio CM, Wong FS, Schloot NC and T-Cell Workshop Committee IoDS (2011). Isolation and preservation of peripheral blood mononuclear cells for analysis of islet antigen-reactive T cell responses: position statement of the T-Cell Workshop Committee of the Immunology of Diabetes Society. *Clin Exp Immunol.* 163(1):33-49.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA and Visscher PM (2009). Finding the missing heritability of complex diseases. *Nature.* 461(7265):747-53.

Mao Y, Mohan R, Zhang S and Tang X (2013). MicroRNAs as pharmacological targets in diabetes. *Pharmacol Res.* 75(37-47.

Marino M, Latrofa F, Menconi F, Chiovato L and Vitti P (2014). Role of genetic and non-genetic factors in the etiology of Graves' disease. *J Endocrinol Invest.*

Martin S, Wolf-Eichbaum D, Duinkerken G, Scherbaum WA, Kolb H, Noordzij JG and Roep BO (2001). Development of type 1 diabetes despite severe hereditary B-lymphocyte deficiency. *N Engl J Med.* 345(14):1036-40.

Martino DJ, Tulic MK, Gordon L, Hodder M, Richman TR, Metcalfe J, Prescott SL and Saffery R (2011). Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics.* 6(9):1085-94.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R and Stamatoyannopoulos JA (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 337(6099):1190-5.

McClay JL, Aberg KA, Clark SL, Nerella S, Kumar G, Xie LY, Hudson AD, Harada A, Hultman CM, Magnusson PK, Sullivan PF and Van Den Oord EJ (2014). A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Hum Mol Genet.* 23(5):1175-85.

McClelland AD and Kantharidis P (2014). microRNA in the development of diabetic complications. *Clin Sci (Lond).* 126(2):95-110.

Medzhitov R (2001). Toll-like receptors and innate immunity. *Nat Rev Immunol.* 1(2):135-45.

Medzhitov R (2007). Recognition of microorganisms and activation of the immune response. *Nature.* 449(7164):819-26.

Medzhitov R and Janeway CA, Jr. (2002). Decoding the patterns of self and nonself by the innate immune system. *Science.* 296(5566):298-300.

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES and Jaenisch R (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33(18):5868-77.

Melloul D, Marshak S and Cerasi E (2002). Regulation of insulin gene transcription. *Diabetologia.* 45(3):309-26.

Merger SR, Leslie RD and Boehm BO (2013). The broad clinical phenotype of Type 1 diabetes at presentation. *Diabet Med.* 30(2):170-8.

Meron MK, Amital H, Shepshelovich D, Barzilai O, Ram M, Anaya JM, Gerli R, Nicola B and Shoenfeld Y (2010). Infectious aspects and the etiopathogenesis of rheumatoid arthritis. *Clin Rev Allergy Immunol.* 38(2-3):287-91.

Metcalfe KA, Hitman GA, Rowe RE, Hawa M, Huang X, Stewart T and Leslie RD (2001). Concordance for type 1 diabetes in identical twins is affected by insulin genotype. *Diabetes Care.* 24(5):838-42.

Miao F, Chen Z, Genuth S, Paterson A, Zhang L, Wu X, Li SM, Cleary P, Riggs A, Harlan DM, Lorenzi G, Kolterman O, Sun W, Lachin JM, Natarajan R and Group DER (2014). Evaluating the role of epigenetic histone modifications in the metabolic memory of type 1 diabetes. *Diabetes.* 63(5):1748-62.

Miao F, Chen Z, Zhang L, Liu Z, Wu X, Yuan YC and Natarajan R (2012). Profiles of epigenetic histone post-translational modifications at type 1 diabetes susceptible genes. *J Biol Chem.* 287(20):16335-45.

Miao F, Smith DD, Zhang L, Min A, Feng W and Natarajan R (2008). Lymphocytes from patients with type 1 diabetes display a distinct profile of chromatin histone H3 lysine 9 dimethylation: an epigenetic study in diabetes. *Diabetes.* 57(12):3189-98.

Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I, Houseman EA, Izzi B, Kelsey KT, Meissner A, Milosavljevic A, Siegmund KD, Bock C and Irizarry RA (2013). Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods.* 10(10):949-55.

Miettinen ME, Reinert L, Kinnunen L, Harjutsalo V, Koskela P, Surcel HM, Lamberg-Allardt C and Tuomilehto J (2012). Serum 25-hydroxyvitamin D level during early pregnancy and type 1 diabetes risk in the offspring. *Diabetologia.* 55(5):1291-4.

Moltchanova EV, Schreier N, Lammi N and Karvonen M (2009). Seasonal variation of diagnosis of Type 1 diabetes mellitus in children worldwide. *Diabet Med.* 26(7):673-8.

Morris TJ and Beck S (2014). Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods.*

Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK and Beck S (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics.* 30(3):428-30.

Mueller SN, Gebhardt T, Carbone FR and Heath WR (2013). Memory T cell subsets, migration patterns, and tissue residence. *Annu Rev Immunol.* 31(137-61.

Nairn C, Galbraith DN, Taylor KW and Clements GB (1999). Enterovirus variants in the serum of children at the onset of Type 1 diabetes mellitus. *Diabet Med.* 16(6):509-13.

Naito T, Tanaka H, Naoe Y and Taniuchi I (2011). Transcriptional control of T-cell development. *Int Immunol.* 23(11):661-8.

Nejentsev S, Howson JM, Walker NM, Szeszko J, Field SF, Stevens HE, Reynolds P, Hardy M, King E, Masters J, Hulme J, Maier LM, Smyth D, Bailey R, Cooper JD, Ribas G, Campbell RD, Clayton DG, Todd JA and Wellcome Trust Case Control C

(2007). Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature.* 450(7171):887-92.

Nejentsev S, Walker N, Riches D, Egholm M and Todd JA (2009). Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science.* 324(5925):387-9.

Nielsen LB, Wang C, Sorensen K, Bang-Berthelsen CH, Hansen L, Andersen ML, Hougaard P, Juul A, Zhang CY, Pociot F and Mortensen HB (2012). Circulating levels of microRNA from children with newly diagnosed type 1 diabetes and healthy controls: evidence that miR-25 associates to residual beta-cell function and glycaemic control during disease progression. *Exp Diabetes Res.* 2012(896362.

Nisticò L, Giorgi G, Giordano M, Galgani A, Petrone A, D'Alfonso S, Federici M, Di Mario U, Pozzilli P, Buzzetti R and Cascino I (2002). IL12B polymorphism and type 1 diabetes in the Italian population: a case-control study. *Diabetes.* 51(5):1649-50.

Noble JA, Valdes AM, Cook M, Klitz W, Thomson G and Erlich HA (1996). The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet.* 59(5):1134-48.

Noble JA, Valdes AM, Varney MD, Carlson JA, Moonsamy P, Fear AL, Lane JA, Lavant E, Rappner R, Louey A, Concannon P, Mychaleckyj JC, Erlich HA and Type 1 Diabetes Genetics C (2010). HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. *Diabetes.* 59(11):2972-9.

Nokoff N and Rewers M (2013). Pathogenesis of type 1 diabetes: lessons from natural history studies of high-risk individuals. *Ann N Y Acad Sci.* 1281(1-15.

Notkins AL and Lernmark A (2001). Autoimmune type 1 diabetes: resolved and unresolved issues. *J Clin Invest.* 108(9):1247-52.

Ochoa-Reparaz J and Kasper LH (2014). Gut microbiome and the risk factors in central nervous system autoimmunity. *FEBS Lett.* 588(22):4214-4222.

Oikarinen M, Tauriainen S, Honkanen T, Oikarinen S, Vuori K, Kaukinen K, Rantala I, Maki M and Hyoty H (2008). Detection of enteroviruses in the intestine of type 1 diabetic patients. *Clin Exp Immunol.* 151(1):71-5.

Oikarinen S, Martiskainen M, Tauriainen S, Huhtala H, Ilonen J, Veijola R, Simell O, Knip M and Hyoty H (2011). Enterovirus RNA in blood is linked to the development of type 1 diabetes. *Diabetes.* 60(1):276-9.

Onkamo P, Vaananen S, Karvonen M and Tuomilehto J (1999). Worldwide increase in incidence of Type I diabetes--the analysis of the data on published incidence trends. *Diabetologia.* 42(12):1395-403.

Orban T, Kis J, Szereday L, Engelmann P, Farkas K, Jalahej H and Treszl A (2007). Reduced CD4+ T-cell-specific gene expression in human type 1 diabetes mellitus. *J Autoimmun.* 28(4):177-87.

Oresic M, Simell S, Sysi-Aho M, Nanto-Salonen K, Seppanen-Laakso T, Parikka V, Katajamaa M, Hekkala A, Mattila I, Keskinen P, Yetukuri L, Reinikainen A, Lahde J, Suortti T, Hakalax J, Simell T, Hyoty H, Veijola R, Ilonen J, Lahesmaa R, Knip M and Simell O (2008). Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. *J Exp Med.* 205(13):2975-84.

Padmos RC, Schloot NC, Beyan H, Ruwhof C, Staal FJ, de Ridder D, Aanstoot HJ, Lam-Tse WK, de Wit H, de Herder C, Drexhage RC, Menart B, Leslie RD, Drexhage HA and Consortium L (2008). Distinct monocyte gene-expression profiles in autoimmune diabetes. *Diabetes.* 57(10):2768-73.

Palmer JP, Hampe CS, Chiu H, Goel A and Brooks-Worrell BM (2005). Is latent autoimmune diabetes in adults distinct from type 1 diabetes or just type 1 diabetes at an older age? *Diabetes.* 54 Suppl 2(S62-7.

Parihar A, Eubank TD and Doseff AI (2010). Monocytes and macrophages regulate immunity through dynamic networks of survival and cell death. *J Innate Immun.* 2(3):204-15.

Parikka V, Nanto-Salonen K, Saarinen M, Simell T, Ilonen J, Hyoty H, Veijola R, Knip M and Simell O (2012). Early seroconversion and rapidly increasing autoantibody concentrations predict prepubertal manifestation of type 1 diabetes in children at genetic risk. *Diabetologia.* 55(7):1926-36.

Parker W (2014). The "hygiene hypothesis" for allergic disease is a misnomer. *BMJ.* 348(g5267.

Parkes M, Cortes A, van Heel DA and Brown MA (2013). Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet.* 14(9):661-73.

Patrick SL, Kadohiro JK, Waxman SH, Curb JD, Orchard TJ, Dorman JS, Kuller LH and LaPorte RE (1997). IDDM incidence in a multiracial population. The Hawaii IDDM Registry, 1980-1990. *Diabetes Care.* 20(6):983-7.

Patterson CC, Dahlquist GG, Gyurus E, Green A, Soltesz G and Group ES (2009). Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet.* 373(9680):2027-33.

Patterson CC, Gyurus E, Rosenbauer J, Cinek O, Neu A, Schober E, Parslow RC, Joner G, Svensson J, Castell C, Bingley PJ, Schoenle E, Jarosz-Chobot P, Urbonaite B, Rothe U, Krzisnik C, Ionescu-Tirgoviste C, Weets I, Kocova M, Stipancic G, Samardzic M, de Beaufort CE, Green A, Dahlquist GG and Soltesz G (2012). Trends in childhood type 1 diabetes incidence in Europe during 1989-2008:

evidence of non-uniformity over time in rates of increase. *Diabetologia.* 55(8):2142-7.

Pereira PF, Alfenas Rde C and Araujo RM (2014). Does breastfeeding influence the risk of developing diabetes mellitus in children? A review of current evidence. *J Pediatr (Rio J).* 90(1):7-15.

Pescovitz MD, Greenbaum CJ, Krause-Steinrauf H, Becker DJ, Gitelman SE, Goland R, Gottlieb PA, Marks JB, McGee PF, Moran AM, Raskin P, Rodriguez H, Schatz DA, Wherrett D, Wilson DM, Lachin JM, Skyler JS and Type 1 Diabetes TrialNet Anti CDSG (2009). Rituximab, B-lymphocyte depletion, and preservation of beta-cell function. *N Engl J Med.* 361(22):2143-52.

Peserico A and Simone C (2011). Physical and functional HAT/HDAC interplay regulates protein acetylation balance. *J Biomed Biotechnol.* 2011(371832.

Petronis A (2006). Epigenetics and twins: three variations on the theme. *Trends Genet.* 22(7):347-50.

Petronis A (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature.* 465(7299):721-7.

Pflueger M, Seppanen-Laakso T, Suortti T, Hyotylainen T, Achenbach P, Bonifacio E, Oresic M and Ziegler AG (2011). Age- and islet autoimmunity-associated differences in amino acid and lipid metabolites in children at risk for type 1 diabetes. *Diabetes.* 60(11):2740-7.

Pidsley R, CC YW, Volta M, Lunnon K, Mill J and Schalkwyk LC (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics.* 14(293.

Pieper K, Grimbacher B and Eibel H (2013). B-cell biology and development. *J Allergy Clin Immunol.* 131(4):959-71.

Pinkse GG, Tysma OH, Bergen CA, Kester MG, Ossendorp F, van Veelen PA, Keymeulen B, Pipeleers D, Drijfhout JW and Roep BO (2005). Autoreactive CD8 T cells associated with beta cell destruction in type 1 diabetes. *Proc Natl Acad Sci U S A.* 102(51):18425-30.

Pociot F, Akolkar B, Concannon P, Erlich HA, Julier C, Morahan G, Nierras CR, Todd JA, Rich SS and Nerup J (2010). Genetics of type 1 diabetes: what's next? *Diabetes.* 59(7):1561-71.

Pociot F and McDermott MF (2002). Genetics of type 1 diabetes mellitus. *Genes and Immunity.* 3(5):235-49.

Poirier LA (2002). The effects of diet, genetics and chemicals on toxicity and aberrant DNA methylation: an introduction. *J Nutr.* 132(8 Suppl):2336S-2339S.

Poulsen P, Esteller M, Vaag A and Fraga MF (2007). The epigenetic basis of twin discordance in age-related diseases. *Pediatr Res.* 61(5 Pt 2):38R-42R.

Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, Daunay A, Busato F, Mein CA, Manfras B, Dias KR, Bell CG, Tost J, Boehm BO, Beck S and Leslie RD (2011a). Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet.* 7(9):e1002300.

Rakyan VK, Down TA, Balding DJ and Beck S (2011b). Epigenome-wide association studies for common human diseases. *Nat Rev Genet.* 12(8):529-41.

Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P and Spector TD (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.* 20(4):434-9.

Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, Tomazou EM, Backdahl L, Johnson N, Herberth M, Howe KL, Jackson DK, Miretti MM, Fiegler H, Marioni JC, Birney E, Hubbard TJ, Carter NP, Tavare S and Beck S (2008). An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.* 18(9):1518-29.

Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox AV, Andrews TD, Howe KL, Otto T, Olek A, Fischer J, Gut IG, Berlin K and Beck S (2004). DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.* 2(12):e405.

Ramagopalan SV and Rakyan VK (2013). The promise and challenges of blood spot methylomics. *Epigenetics.* 8(8):775-7.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP and Jaenisch R (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A.* 97(10):5237-42.

Raymond NT, Jones JR, Swift PG, Davies MJ, Lawrence G, McNally PG, Burden ML, Gregory R, Burden AC and Botha JL (2001). Comparative incidence of Type I diabetes in children aged under 15 years from South Asian and White or Other ethnic backgrounds in Leicestershire, UK, 1989 to 1998. *Diabetologia.* 44 Suppl 3(B32-6.

Redondo MJ, Jeffrey J, Fain PR, Eisenbarth GS and Orban T (2008). Concordance for islet autoimmunity among monozygotic twins. *N Engl J Med.* 359(26):2849-50.

Redondo MJ, Rewers M, Yu L, Garg S, Pilcher CC, Elliott RB and Eisenbarth GS (1999). Genetic determination of islet cell autoimmunity in monozygotic twin, dizygotic twin, and non-twin siblings of patients with type 1 diabetes: prospective twin study. *BMJ.* 318(7185):698-702.

Redondo MJ, Yu L, Hawa M, Mackenzie T, Pyke DA, Eisenbarth GS and Leslie RD (2001). Heterogeneity of type I diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia.* 44(3):354-62.

Reik W, Dean W and Walter J (2001). Epigenetic reprogramming in mammalian development. *Science.* 293(5532):1089-93.

Reik W and Walter J (2001). Genomic imprinting: parental influence on the genome. *Nat Rev Genet.* 2(1):21-32.

Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, Soderhall C, Scheynius A and Kere J (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One.* 7(7):e41361.

Renz H, von Mutius E, Brandtzaeg P, Cookson WO, Autenrieth IB and Haller D (2011). Gene-environment interactions in chronic inflammatory disease. *Nat Immunol.* 12(4):273-7.

Rhodes A, Wort SJ, Thomas H, Collinson P and Bennett ED (2006). Plasma DNA concentration as a predictor of mortality and sepsis in critically ill patients. *Crit Care.* 10(2):R60.

Ringner M (2008). What is principal component analysis? *Nat Biotechnol.* 26(3):303-4.

Rioux JD and Abbas AK (2005). Paths to understanding the genetic basis of autoimmune disease. *Nature.* 435(7042):584-9.

Rodriguez-Paredes M and Esteller M (2011). Cancer epigenetics reaches mainstream oncology. *Nat Med.* 17(3):330-9.

Roep BO (2003). The role of T-cells in the pathogenesis of Type 1 diabetes: from cause to cure. *Diabetologia.* 46(3):305-21.

Roep BO and Peakman M (2011). Diabetogenic T lymphocytes in human Type 1 diabetes. *Curr Opin Immunol.* 23(6):746-53.

Roggli E, Gattesco S, Caille D, Briet C, Boitard C, Meda P and Regazzi R (2012). Changes in microRNA expression contribute to pancreatic beta-cell dysfunction in prediabetic NOD mice. *Diabetes.* 61(7):1742-51.

Rose NR and Bona C (1993). Defining criteria for autoimmune diseases (Witebsky's postulates revisited). *Immunol Today.* 14(9):426-30.

Rowe JH, Ertelt JM and Way SS (2012). Foxp3(+) regulatory T cells, immune stimulation and host defence against infection. *Immunology.* 136(1):1-10.

Ruan Q, Wang T, Kameswaran V, Wei Q, Johnson DS, Matschinsky F, Shi W and Chen YH (2011). The microRNA-21-PDCD4 axis prevents type 1 diabetes by blocking pancreatic beta cell death. *Proc Natl Acad Sci U S A.* 108(29):12030-5.

Ruiz-Esquide V and Sanmarti R (2012). Tobacco and other environmental risk factors in rheumatoid arthritis. *Reumatol Clin.* 8(6):342-50.

Sadeharju K, Knip M, Hiltunen M, Akerblom HK and Hyoty H (2003). The HLA-DR phenotype modulates the humoral immune response to enterovirus antigens. *Diabetologia.* 46(8):1100-5.

Saha P and Geissmann F (2011). Toward a functional characterization of blood monocytes. *Immunol Cell Biol.* 89(1):2-4.

Sakaguchi S, Yamaguchi T, Nomura T and Ono M (2008). Regulatory T cells and immune tolerance. *Cell.* 133(5):775-87.

Salvatoni A, Baj A, Bianchi G, Federico G, Colombo M and Toniolo A (2013). Intrafamilial spread of enterovirus infections at the clinical onset of type 1 diabetes. *Pediatr Diabetes.* 14(6):407-16.

Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M and Esteller M (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics.* 6(6):692-702.

Santin I, Castellanos-Rubio A, Aransay AM, Gutierrez G, Gaztambide S, Rica I, Vicario JL, Noble JA, Castano L and Bilbao JR (2009). Exploring the diabetogenicity of the HLA-B18-DR3 CEH: independent association with T1D genetic risk close to HLA-DOA. *Genes and Immunity.* 10(6):596-600.

Santin I, Moore F, Colli ML, Gurzov EN, Marselli L, Marchetti P and Eizirik DL (2011). PTPN2, a candidate gene for type 1 diabetes, modulates pancreatic beta-cell apoptosis via regulation of the BH3-only protein Bim. *Diabetes.* 60(12):3279-88.

Sanyal A, Lajoie BR, Jain G and Dekker J (2012). The long-range interaction landscape of gene promoters. *Nature.* 489(7414):109-13.

Schenten D and Medzhitov R (2011). The control of adaptive immune responses by the innate immune system. *Adv Immunol.* 109(87-124.

Schlosser M, Mueller PW, Torn C, Bonifacio E, Bingley PJ and Participating L (2010). Diabetes Antibody Standardization Program: evaluation of assays for insulin autoantibodies. *Diabetologia.* 53(12):2611-20.

Sebastiani G, Grieco FA, Spagnuolo I, Galleri L, Cataldo D and Dotta F (2011). Increased expression of microRNA miR-326 in type 1 diabetic patients with ongoing islet autoimmunity. *Diabetes Metab Res Rev.* 27(8):862-6.

Seiskari T, Kondrashova A, Viskari H, Kaila M, Haapala AM, Aittoniemi J, Virta M, Hurme M, Uibo R, Knip M, Hyoty H and group Es (2007). Allergic sensitization and microbial load--a comparison between Finland and Russian Karelia. *Clin Exp Immunol.* 148(1):47-52.

Serrano-Rios M, Goday A and Martinez Larrad T (1999). Migrant populations and the incidence of type 1 diabetes mellitus: an overview of the literature with a focus on the Spanish-heritage countries in Latin America. *Diabetes Metab Res Rev.* 15(2):113-32.

Sharp AJ, Stathaki E, Migliavacca E, Brahmachary M, Montgomery SB, Dupre Y and Antonarakis SE (2011). DNA methylation profiles of human active and inactive X chromosomes. *Genome Res.* 21(10):1592-600.

Siljander HT, Simell S, Hekkala A, Lahde J, Simell T, Vahasalo P, Veijola R, Ilonen J, Simell O and Knip M (2009). Predictive characteristics of diabetes-associated autoantibodies among children with HLA-conferred disease susceptibility in the general population. *Diabetes.* 58(12):2835-42.

Silva DG, Daley SR, Hogan J, Lee SK, Teh CE, Hu DY, Lam KP, Goodnow CC and Vinuesa CG (2011). Anti-islet autoantibodies trigger autoimmune diabetes in the presence of an increased frequency of islet-reactive CD4 T cells. *Diabetes.* 60(8):2102-11.

Simonen-Tikka ML, Pflueger M, Klemola P, Savolainen-Kopra C, Smura T, Hummel S, Kaijalainen S, Nuutila K, Natri O, Roivainen M and Ziegler AG (2011). Human enterovirus infections in children at increased risk for type 1 diabetes: the Babydiet study. *Diabetologia.* 54(12):2995-3002.

Simpson M, Brady H, Yin X, Seifert J, Barriga K, Hoffman M, Bugawan T, Baron AE, Sokol RJ, Eisenbarth G, Erlich H, Rewers M and Norris JM (2011). No association of vitamin D intake or 25-hydroxyvitamin D levels in childhood with risk of islet autoimmunity and type 1 diabetes: the Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia.* 54(11):2779-88.

Siqueira JF, Jr., Fouad AF and Rocas IN (2012). Pyrosequencing as a tool for better understanding of human microbiomes. *J Oral Microbiol.* 4(

Slatkin M (2009). Epigenetic inheritance and the missing heritability problem. *Genetics.* 182(3):845-50.

Smith ZD and Meissner A (2013). DNA methylation: roles in mammalian development. *Nat Rev Genet.* 14(3):204-20.

Soderstrom U, Aman J and Hjern A (2012). Being born in Sweden increases the risk for type 1 diabetes - a study of migration of children to Sweden as a natural experiment. *Acta Paediatr.* 101(1):73-7.

Soltesz G, Patterson CC, Dahlquist G and Group ES (2007). Worldwide childhood type 1 diabetes incidence--what can we learn from epidemiology? *Pediatr Diabetes.* 8 Suppl 6(6-14.

Sørensen IM, Joner G, Jenum PA, Eskild A, Torjesen PA and Stene LC (2012). Maternal serum levels of 25-hydroxy-vitamin d during pregnancy and risk of type 1 diabetes in the offspring. *Diabetes.* 61(1):175-8.

Soriano-Tarraga C, Jimenez-Conde J, Giralt-Steinhauer E, Ois A, Rodriguez-Campello A, Cuadrado-Godia E, Fernandez-Cadenas I, Montaner J, Lucas G, Elosua R, Roquer J and GeneStroke "The Spanish Stroke Genetics C (2013). DNA isolation method is a source of global DNA methylation variability measured with LUMA. Experimental analysis and a systematic review. *PLoS One.* 8(4):e60750.

Steck AK and Rewers MJ (2011). Genetics of type 1 diabetes. *Clin Chem.* 57(2):176-85.

Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, Teschendorff AE, Butcher LM, Beck S and Muller M (2014). Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr).* 36(3):9648.

Stefan M, Zhang W, Concepcion E, Yi Z and Tomer Y (2013). DNA methylation profiles in type 1 diabetes twins point to strong epigenetic effects on etiology. *J Autoimmun.*

Stene LC, Honeyman MC, Hoffenberg EJ, Haas JE, Sokol RJ, Emery L, Taki I, Norris JM, Erlich HA, Eisenbarth GS and Rewers M (2006). Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study. *Am J Gastroenterol.* 101(10):2333-40.

Stene LC, Oikarinen S, Hyoty H, Barriga KJ, Norris JM, Klingensmith G, Hutton JC, Erlich HA, Eisenbarth GS and Rewers M (2010). Enterovirus infection and progression from islet autoimmunity to type 1 diabetes: the Diabetes and Autoimmunity Study in the Young (DAISY). *Diabetes.* 59(12):3174-80.

Stenstrom G, Gottsater A, Bakhtadze E, Berger B and Sundkvist G (2005). Latent autoimmune diabetes in adults: definition, prevalence, beta-cell function, and treatment. *Diabetes.* 54 Suppl 2(S68-72.

Steves CJ, Spector TD and Jackson SH (2012). Ageing, genes, environment and epigenetics: what twin studies tell us now, and in the future. *Age Ageing.* 41(5):581-6.

Stewart SK, Morris TJ, Guilhamon P, Bulstrode H, Bachman M, Balasubramanian S and Beck S (2014). oxBS-450K: A method for analysing hydroxymethylation using 450K BeadChips. *Methods.*

Strahl BD and Allis CD (2000). The language of covalent histone modifications. *Nature.* 403(6765):41-5.

Sumpter KM, Adhikari S, Grishman EK and White PC (2011). Preliminary studies related to anti-interleukin-1beta therapy in children with newly diagnosed type 1 diabetes. *Pediatr Diabetes.* 12(7):656-67.

Swarup V and Rajeswari MR (2007). Circulating (cell-free) nucleic acids--a promising, non-invasive tool for early detection of several human diseases. *FEBS Lett.* 581(5):795-9.

Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, Beck S and Butcher LM (2012). Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat Protoc.* 7(4):617-36.

Talbert PB and Henikoff S (2006). Spreading of silent chromatin: inaction at a distance. *Nat Rev Genet.* 7(10):793-803.

Tanaka K and Okamoto A (2007). Degradation of DNA by bisulfite treatment. *Bioorg Med Chem Lett.* 17(7):1912-5.

Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D and Beck S (2013a). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics.* 29(2):189-96.

Teschendorff AE, West J and Beck S (2013b). Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum Mol Genet.* 22(R1):R7-R15.

Thirlwell C, Eymard M, Feber A, Teschendorff A, Pearce K, Lechner M, Widschwendter M and Beck S (2010). Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. *Methods.* 52(3):248-54.

Thompson RF, Atzmon G, Gheorghe C, Liang HQ, Lowes C, Greally JM and Barzilai N (2010). Tissue-specific dysregulation of DNA methylation in aging. *Aging Cell.* 9(4):506-18.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE and Stamatoyannopoulos JA (2012). The accessible chromatin landscape of the human genome. *Nature.* 489(7414):75-82.

Tikoo K, Meena RL, Kabra DG and Gaikwad AB (2008). Change in post-translational modifications of histone H3, heat-shock protein-27 and MAP kinase p38 expression by curcumin in streptozotocin-induced type I diabetic nephropathy. *Br J Pharmacol.* 153(6):1225-31.

Tobon GJ, Youinou P and Saraux A (2010). The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis. *Autoimmun Rev.* 9(5):A288-92.

Todd JA (2010). Etiology of type 1 diabetes. *Immunity.* 32(4):457-67.

Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, Wallace C, Howson JM, Guja C, Ionescu-Tîrgovişte C, Simmonds MJ, Heward JM, Gough SC, Dunger DB, Wicker LS, Clayton DG, Finland GoTDi and Consortium WTCC (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet.* 39(7):857-64.

Tomankova T, Petrek M, Gallo J and Kriegova E (2011). MicroRNAs: emerging regulators of immune-mediated diseases. *Scand J Immunol.*

Tong YK and Lo YM (2006). Diagnostic developments involving cell-free (circulating) nucleic acids. *Clin Chim Acta.* 363(1-2):187-96.

Touleimat N and Tost J (2012). Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics.* 4(3):325-41.

Troncone R and Jabri B (2011). Coeliac disease and gluten sensitivity. *J Intern Med.* 269(6):582-90.

Tserel L, Kolde R, Rebane A, Kisand K, Org T, Peterson H, Vilo J and Peterson P (2010). Genome-wide promoter analysis of histone modifications in human monocyte-derived antigen presenting cells. *BMC Genomics.* 11(642.

Turvey SE and Broide DH (2010). Innate immunity. *J Allergy Clin Immunol.* 125(2 Suppl 2):S24-32.

Umetani N, Hiramatsu S and Hoon DS (2006). Higher amount of free circulating DNA in serum than in plasma is not mainly caused by contaminated extraneous DNA during separation. *Ann N Y Acad Sci.* 1075(299-307.

Usmani-Brown S, Lebastchi J, Steck AK, Beam C, Herold KC and Ledizet M (2014). Analysis of beta-Cell Death in Type 1 Diabetes by Droplet Digital PCR. *Endocrinology.* 155(9):3694-8.

Vaarala O (2013). Human intestinal microbiota and type 1 diabetes. *Curr Diab Rep.* 13(5):601-7.

Vaarala O, Klemetti P, Juhela S, Simell O, Hyoty H and Ilonen J (2002). Effect of coincident enterovirus infection and cows' milk exposure on immunisation to insulin in early infancy. *Diabetologia.* 45(4):531-4.

Vang T, Congia M, Macis MD, Musumeci L, Orru V, Zavattari P, Nika K, Tautz L, Tasken K, Cucca F, Mustelin T and Bottini N (2005). Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet.* 37(12):1317-9.

Vatay A, Rajczy K, Pozsonyi E, Hosszufalusi N, Prohaszka Z, Fust G, Karadi I, Szalai C, Grosz A, Bartfai Z and Panczel P (2002). Differences in the genetic background of latent autoimmune diabetes in adults (LADA) and type 1 diabetes mellitus. *Immunol Lett.* 84(2):109-115.

Virtanen SM, Laara E, Hypponen E, Reijonen H, Rasanen L, Aro A, Knip M, Ilonen J and Akerblom HK (2000). Cow's milk consumption, HLA-DQB1 genotype, and type 1 diabetes: a nested case-control study of siblings of children with diabetes. Childhood diabetes in Finland study group. *Diabetes.* 49(6):912-7.

Wahren J, Kallas A and Sima AA (2012). The clinical potential of C-peptide replacement in type 1 diabetes. *Diabetes.* 61(4):761-72.

Wang J, Liu L, Ma J, Sun F, Zhao Z and Gu M (2014). Common variants on cytotoxic T lymphocyte antigen-4 polymorphisms contributes to type 1 diabetes susceptibility: evidence based on 58 studies. *PLoS One.* 9(1):e85982.

Wang L, Lovejoy NF and Faustman DL (2012). Persistence of prolonged C-peptide production in type 1 diabetes as measured with an ultrasensitive C-peptide assay. *Diabetes Care.* 35(3):465-70.

Wellcome Trust Case Control C (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 447(7145):661-78.

Wenzlau JM and Hutton JC (2013). Novel diabetes autoantibodies and prediction of type 1 diabetes. *Curr Diab Rep.* 13(5):608-15.

Wenzlau JM, Juhl K, Yu L, Moua O, Sarkar SA, Gottlieb P, Rewers M, Eisenbarth GS, Jensen J, Davidson HW and Hutton JC (2007). The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc Natl Acad Sci U S A.* 104(43):17040-5.

Wielscher M, Pulverer W, Peham J, Hofner M, Rappaport CF, Singer C, Jungbauer C, Nohammer C and Weinhausel A (2011). Methyl-binding domain protein-based DNA isolation from human blood serum combines DNA analyses and serum-autoantibody testing. *BMC Clin Pathol.* 11(11.

Wiersinga WM (2013). Smoking and thyroid. *Clin Endocrinol (Oxf).* 79(2):145-51.

Wild S, Roglic G, Green A, Sicree R and King H (2004). Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care.* 27(5):1047-53.

Wilhelm-Benartzi CS, Koestler DC, Karagas MR, Flanagan JM, Christensen BC, Kelsey KT, Marsit CJ, Houseman EA and Brown R (2013). Review of processing and analysis methods for DNA methylation array data. *Br J Cancer.* 109(6):1394-402.

Willcox A, Richardson SJ, Bone AJ, Foulis AK and Morgan NG (2009). Analysis of islet inflammation in human type 1 diabetes. *Clin Exp Immunol.* 155(2):173-81.

Winkler C, Lauber C, Adler K, Grallert H, Illig T, Ziegler AG and Bonifacio E (2011). An interferon-induced helicase (IFIH1) gene polymorphism associates with different rates of progression from autoimmunity to type 1 diabetes. *Diabetes.* 60(2):685-90.

Winter WE and Schatz DA (2011). Autoimmune markers in diabetes. *Clin Chem.* 57(2):168-75.

Wong FS, Wen L, Tang M, Ramanathan M, Visintin I, Daugherty J, Hannum LG, Janeway CA, Jr. and Shlomchik MJ (2004). Investigation of the role of B-cells in type 1 diabetes in the NOD mouse. *Diabetes.* 53(10):2581-7.

Wong KL, Yeap WH, Tai JJ, Ong SM, Dang TM and Wong SC (2012). The three human monocyte subsets: implications for health and disease. *Immunol Res.* 53(1-3):41-57.

Xiao X, Chen Z, Shiota C, Prasadan K, Guo P, El-Gohary Y, Paredes J, Welsh C, Wiersch J and Gittes GK (2013). No evidence for beta cell neogenesis in murine adult pancreas. *J Clin Invest.* 123(5):2207-17.

Yang C, Chapman AG, Kelsey AD, Minks J, Cotton AM and Brown CJ (2011). X-chromosome inactivation: molecular mechanisms from the human perspective. *Hum Genet.* 130(2):175-85.

Yeung WC, Rawlinson WD and Craig ME (2011). Enterovirus infection and type 1 diabetes mellitus: systematic review and meta-analysis of observational molecular studies. *BMJ.* 342(d35.

Zaratiegui M, Irvine DV and Martienssen RA (2007). Noncoding RNAs and gene silencing. *Cell.* 128(4):763-76.

Zenewicz LA, Abraham C, Flavell RA and Cho JH (2010). Unraveling the genetics of autoimmunity. *Cell.* 140(6):791-7.

Zhang H, Liao X, Sparks JB and Luo XM (2014). Dynamics of gut microbiota in autoimmune lupus. *Appl Environ Microbiol.* 80(24):7551-60.

Zhao C, Tan YC, Wong WC, Sem X, Zhang H, Han H, Ong SM, Wong KL, Yeap WH, Sze SK, Kourilsky P and Wong SC (2010). The CD14(+/low)CD16(+) monocyte subset is more susceptible to spontaneous and oxidant-induced apoptosis than the CD14(+)CD16(-) subset. *Cell Death Dis.* 1(e95.

Zhao E, Xu H, Wang L, Kryczek I, Wu K, Hu Y, Wang G and Zou W (2012). Bone marrow and the control of immunity. *Cell Mol Immunol.* 9(1):11-9.

Zhao Z and Han L (2009). CpG islands: algorithms and applications in methylation studies. *Biochem Biophys Res Commun.* 382(4):643-5.

Zhou L, Somasundaram R, Nederhof RF, Dijkstra G, Faber KN, Peppelenbosch MP and Fuhler GM (2012). Impact of human granulocyte and monocyte isolation procedures on functional studies. *Clin Vaccine Immunol.* 19(7):1065-74.

Zhou Z and Jensen PE (2013). Structural Characteristics of HLA-DQ that May Impact DM Editing and Susceptibility to Type-1 Diabetes. *Front Immunol.* 4(262.

Zhou Z, Xiang Y, Ji L, Jia W, Ning G, Huang G, Yang L, Lin J, Liu Z, Hagopian WA, Leslie RD and Group LCS (2013). Frequency, immunogenetics, and clinical characteristics of latent autoimmune diabetes in China (LADA China study): a nationwide, multicenter, clinic-based cross-sectional study. *Diabetes.* 62(2):543-50.

Zhu ZZ, Hou L, Bollati V, Tarantini L, Marinelli B, Cantone L, Yang AS, Vokonas P, Lissowska J, Fustinoni S, Pesatori AC, Bonzini M, Apostoli P, Costa G, Bertazzi PA, Chow WH, Schwartz J and Baccarelli A (2012). Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. *Int J Epidemiol.* 41(1):126-39.

Ziegler A, Zangemeister-Wittke U and Stahel RA (2002). Circulating DNA: a new diagnostic gold mine? *Cancer Treat Rev.* 28(5):255-71.

Ziegler AG, Bonifacio E and Group B-BS (2012). Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. *Diabetologia.* 55(7):1937-43.

Ziegler AG and Nepom GT (2010). Prediction and pathogenesis in type 1 diabetes. *Immunity.* 32(4):468-78.

Ziegler AG, Schmid S, Huber D, Hummel M and Bonifacio E (2003). Early infant feeding and risk of developing type 1 diabetes-associated autoantibodies. *JAMA.* 290(13):1721-8.

Ziegler AI, Le Page MA, Maxwell MJ, Stolp J, Guo H, Jayasimhan A, Hibbs ML, Santamaria P, Miller JF, Plebanski M, Silveira PA and Slattery RM (2013). The CD19 signalling molecule is elevated in NOD mice and controls type 1 diabetes development. *Diabetologia.* 56(12):2659-68.

Zilbauer M, Rayner TF, Clark C, Coffey AJ, Joyce CJ, Palta P, Palotie A, Lyons PA and Smith KG (2013). Genome-wide methylation analyses of primary human leukocyte subsets identifies functionally important cell-type-specific hypomethylated regions. *Blood.* 122(25):e52-60.

Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A and Meissner A (2013). Charting a

dynamic DNA methylation landscape of the human genome. *Nature.* 500(7463):477-81.

Zipper H, Brunner H, Bernhagen J and Vitzthum F (2004). Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Res.* 32(12):e103.

Zipris D (2009). Epidemiology of type 1 diabetes and what animal models teach us about the role of viruses in disease mechanisms. *Clin Immunol.* 131(1):11-23.

Chapter 8

Appendix

8 Appendix

8.1 Appendix I – Materials and equipment

8.1.1 Chemical reagents and enzymes

10x PfuTurbo Cx reaction buffer	Agilent Technologies, UK
20% Human Albumin	PAA, UK
2-Mercaptoethanol	Sigma Aldrich, UK
2-propanol	Sigma Aldrich, UK
50 bp ladder	Invitrogen, UK
Agencourt AMPure XP - PCR Purification	Beckman Coulter, UK
Ammonium chloride	Sigma Aldrich, UK
Ampicillin	Sigma Aldrich, UK
BglII	New England Biolabs, UK
BamHI	New England Biolabs, UK
Boric acid	Sigma Aldrich, UK
CpG Methyltransferase (M.SssI)	New England Biolabs, UK
dATP PCR Grade, sodium salt	Roche Applied Science, UK
dNTP mix	New England Biolabs, UK
DEPC water, molecular biology grade	Invitrogen, UK
DNA Polymerase I, Large (Klenow) Fragment	New England Biolabs, UK
EDTA	Sigma Aldrich, UK
Ethanol	Sigma Aldrich, UK
Ethidium bromide	Sigma Aldrich, UK
Klenow Fragment (3′→5′ exo-)	New England Biolabs, UK
LB Agar, powder (Lennox)	Life technologies, UK
LB Broth, powder (Lennox)	Life technologies, UK
MESA BLUE qPCR MasterMix Plus for SYBR® Assay	Eurogentec, UK
MyTaq™ HS DNA Polymerase	Bioline, UK
NEBuffer 3	New England Biolabs, UK
NotI	New England Biolabs, UK
PBS 1X	Sigma Aldrich, UK
PBS 10X	Sigma Aldrich, UK
Percoll	GE Healthcare, UK
PfuTurbo Cx Hotstart DNA Polymerase	Agilent Technologies, UK

Penicillin 1000u/ml-Streptomycin100µg/ml	Invitrogen, UK
Potassium bicarbonate	Sigma Aldrich, UK
Power SYBR® Green PCR Master Mix	Life Technologies, UK
Protein A Sepharose	GE Healthcare, UK
QuantiTect SYBR Green PCR Kit	Qiagen, UK
Quick Ligation™ Kit	New England Biolabs, UK
Scintillation fluid	PerkinElmer, UK
Sodium citrate tribasic dihydrate	Sigma Aldrich, UK
T4 DNA Polymerase	New England Biolabs, UK
T4 Polynucleotide Kinase	New England Biolabs, UK
TaqMan® Universal PCR Master Mix	Life Technologies, UK
TaqMan® Gene Expression Master Mix	Life Technologies, UK
Tris-Base	Sigma Aldrich, UK
Trypan blue 0.4%	Sigma Aldrich, UK
UltraPure™ Agarose	Invitrogen, UK

8.1.2 Equipment

7500 Real Time PCR System	Applied Biosystems, UK
Agarose gel electrophoresis tank	Bio-rad, UK
BD FACS Canto II	BD, UK
Bioanalyzer	Agilent Technologies, UK
Biorupter	Diagenode, BE
Glucose meter	HemoCue, UK
Haemocytometer	Hawksley, UK
Heatblock	Grant Instruments, UK
MACS stand and separator	Miltenyi Biotech, UK
Megafuge® 1.0 R	Heraeus, UK
Microcentrifuge	Hettich, SZ
Nanodrop	Thermo Scientific, UK
PTC-225 Peltier Thermal Cycler	MJ Research, US
Pyrosequencer PSQ96	Qiagen, UK
Qubit Fluorometer	Invitrogen, UK
QuadroMACS Separation Unit	Miltenyi Biotech, UK
Roller	Luckham ltd, UK
SafeFAST Elite fume hood	Faster S.r.l., IT
Shandon Cytospin 3 Centrifuge	Thermo Scientific, UK
UV Transilluminator	Alpha Innotech Corporation, US
Vortex	Grant Instruments, UK
Water bath	Grant Instruments, UK

8.1.3 Consumables

50ml and 15ml Flacon tubes	Sarstedt, DE
1.5ml eppendorf tubes	Eppendorf, UK
0.2ml pcr tubes	Starlab, UK
Adhesive PCR sealing film	Invitrogen, UK
96 well flat bottom plates	Orange Scientific, BE
96 well plates	Invitrogen, UK
96-well plates (filtered)	Milipore, UK
BD Vacutainer® Sodium Heparin tube	BD, UK
BD Vacutainer® Plastic Serum tube with Red BD Hemogard™ closure	BD, UK
BD Vacutainer® Plastic K3EDTA tube	BD, UK
B-glucose microcuvettes	HemoCue, UK
BD Safety-Lok; Blood Collection Set	BD, UK
Cryovials	Starlab, UK
Falcon® 5mL Round Bottom Polystyrene Test Tube	Corning, UK
Filter Cards for Shandon* Cytospin	Fisher Scientific, UK
LD columns	Miltenyi Biotech, UK
LS columns	Miltenyi Biotech, UK
MS columns	Miltenyi Biotech, UK
Pre-separation filters	Miltenyi Biotech, UK
Pasteur Pipette Single Wrap	SLS, UK
Slides	Invitrogen, UK

8.1.4 Kits

Agilent High Sensitivity DNA Kit	Agilent Technologies, UK
DNA Clean and Concentrator-5 kit	Zymo Research, US
EpiTect Bisulfite Kit	Qiagen, UK
EZ DNA Methylation™ Kit	Zymo Research, US
EZ-96 DNA Methylation MagPrep kit	Zymo Research, US
GeneJET Plasmid Miniprerp kit	Thermo Scientific, UK
GenSolve DNA recovery kit	Labtech, UK
Gentra Puregene Buccal Cell Kit	Qiagen, UK
Ovation Ultralow Library Systems	Nugen, US
QIAamp DNA Blood Mini Kit	Qiagen, UK
QIAquick Gel Extraction Kit	Qiagen, UK
QIAquick PCR Purification Kit	Qiagen, UK
Qubit® dsDNA HS Assay Kit	Invitrogen, UK
Qubit® dsDNA BR Assay Kit	Invitrogen, UK
Rapid Romanowsky stain kit	Fisher Scientific, UK
RNeasy Mini kit	Qiagen, UK

8.1.5 Buffers and Media

10 xTTBS	78.8 g Tris 116.8 g sodium chloride Adjust to pH 7.2 100 mL of Tween 20
1M trisodium citrate	29.4 g trisodium citrate in 100ml of water through a filter
Buffer 2	PBS 1X 2 mM EDTA 500 mL PBS 2 mL 0.5M EDTA
Buffer 3	PBS 1X 13mM NaCitrate 500 mL PBS 6.6 mL 1 M trisodium citrate
LB agar	6.4 g of LB agar powder topped up to 500 mL with distilled water 100 µL ampicillin
LB broth	8 g of LB broth powder, topped up to 500 mL with distilled water 100 µL ampicillin
Lysis Buffer	4.15 g ammonium chloride 0.5 g potassium bicarbonate 100 µL 0.5 M EDTA to 500 mL of water
Percoll 1.078 g/mL (561 mL)	300 mL Percoll 16 mL PBS 10X 144 mL PBS 1X 9.2 mL 20% human albumin 4.8 mL 1 M trisodium citrate
RPMI	5 mL penicillin streptomycin 5 mL human serum albumin to 500 mL RPMI 1640 Medium, GlutaMAX™, HEPES (Invitrogen)
TBE buffer 10x	108 g tris base 51g boric acid 40 mL EDTA 500 mM pH 8.0

8.1.6 Antibodies

Anti-Mouse Ig, κ/Negative Control (FBS) Compensation Particles Set	BD Biosciences, UK
CD14 microbeads, human	Miltenyi Biotech, UK
CD16 microbeads, human	Miltenyi Biotech, UK
CD4 microbeads, human	Miltenyi Biotech, UK
CD19 FITC, human	Miltenyi Biotech, UK
CD19 microbeads, human	Miltenyi Biotech, UK
CD66b FITC antibody	NHS, UK
CD16 PE antibody	Miltenyi Biotech, UK
CD45 PE antibody	Invitrogen, UK
FITC conjugated monoclonal mouse anti-human CD14, clone MφP9	BD Biosciences, UK
FITC conjugated monoclonal mouse anti-human CD4, clone M-T466	Miltenyi Biotech, UK
Ig, κ light chain FITC TB28-2 Ms IgG1, κ	BD Biosciences, UK
Ig, κ Light Chain PE TB28-2 Ms IgG1, κ	BD Biosciences, UK
PE conjugated monoclonal mouse anti-human CD19, clone LT19	Miltenyi Biotech, UK
PE conjugated monoclonal mouse anti-human CD16, clone B73.1/leu11c	BD Biosciences, UK
PerCP-Cy5.5 conjugated monoclonal mouse anti-human CD64, clone 10.1	BD Biosciences, UK
PE-CY7 conjugated monoclonal mouse anti-human CD45, clone HI30	Invitrogen, UK
PerCP-Cy™5.5 Mouse IgG1 κ Isotype Control	BD Biosciences, UK
PE-Cy® 7 Mouse IgG1	Invitrogen, UK

8.1.7 Databases and software

7500 Software v2.0.6	
Chromas Lite	http://chromas-lite.software.informer.com/
DAVID	http://david.abcc.ncifcrf.gov/tools.jsp
Flowjo v8.0 and v10.0	
GraphPad Prism v5.0	
MethPrimer design	http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi
Microsoft Office 2007 and 2010	
PubMed	http://www.ncbi.nlm.nih.gov/pubmed/
R	http://www.r-project.org/
SPSS v17.0	
UCSC Genome Browser	http://genome.ucsc.edu/

8.1.8 DNA oligonucleotides

Table 8.1. Table of DNA oligonucleotides.

All primers were purchased from Sigma Aldrich, UK.

Bisulfite sequencing	
PE PCR Primer 1.01	5'– AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCT*T
PE PCR Primer 2.01	5'– CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATT CCTGCTGAACCGCTCTTCCGATC*T
	Methylated adapters PE Adapters1 5' P-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Methylation specific PCR	
First step PCR_Forward	TTAGGGGTTTTAAGGTAGGGTATTTGGT
First step PCR_Reverse	ACCAAAAACAACAATAAACAATTAACTCACCCTACAA
MSP_Methylated Forward	TAGTCGTAGTTTTTGTGAATTAATATTTGTGC
MSP_Methylated Reverse	CACCCTACAAATCCTCTACCTCCCG
MSP_Unmethylated Forward	TTAGTTGTAGTTTTTGTGAATTAATATTTGTGT
MSP_Unmethylated Reverse	CACCCTACAAATCCTCTACCTCCCA

8.2 Appendix II - Methods

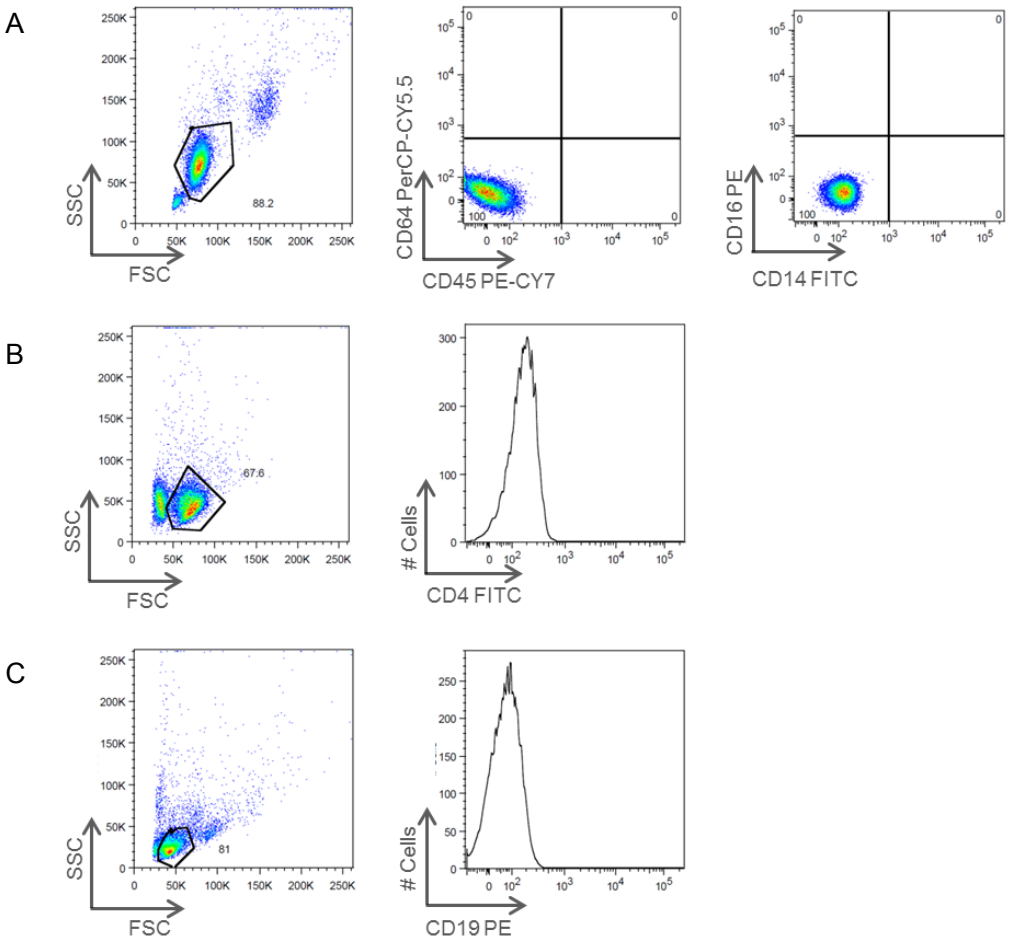


Figure 8.1. Unstained controls for FACS analysis. Unstained controls for each cell type were analysed alongside $CD14^{+}CD16^{-}$ monocytes (A), $CD4^{+}$ T cells (B) and $CD19^{+}$ B cells (C) stained with the conjugated antibodies. Unstained controls were used to detect autofluorescence or background staining of the cells.

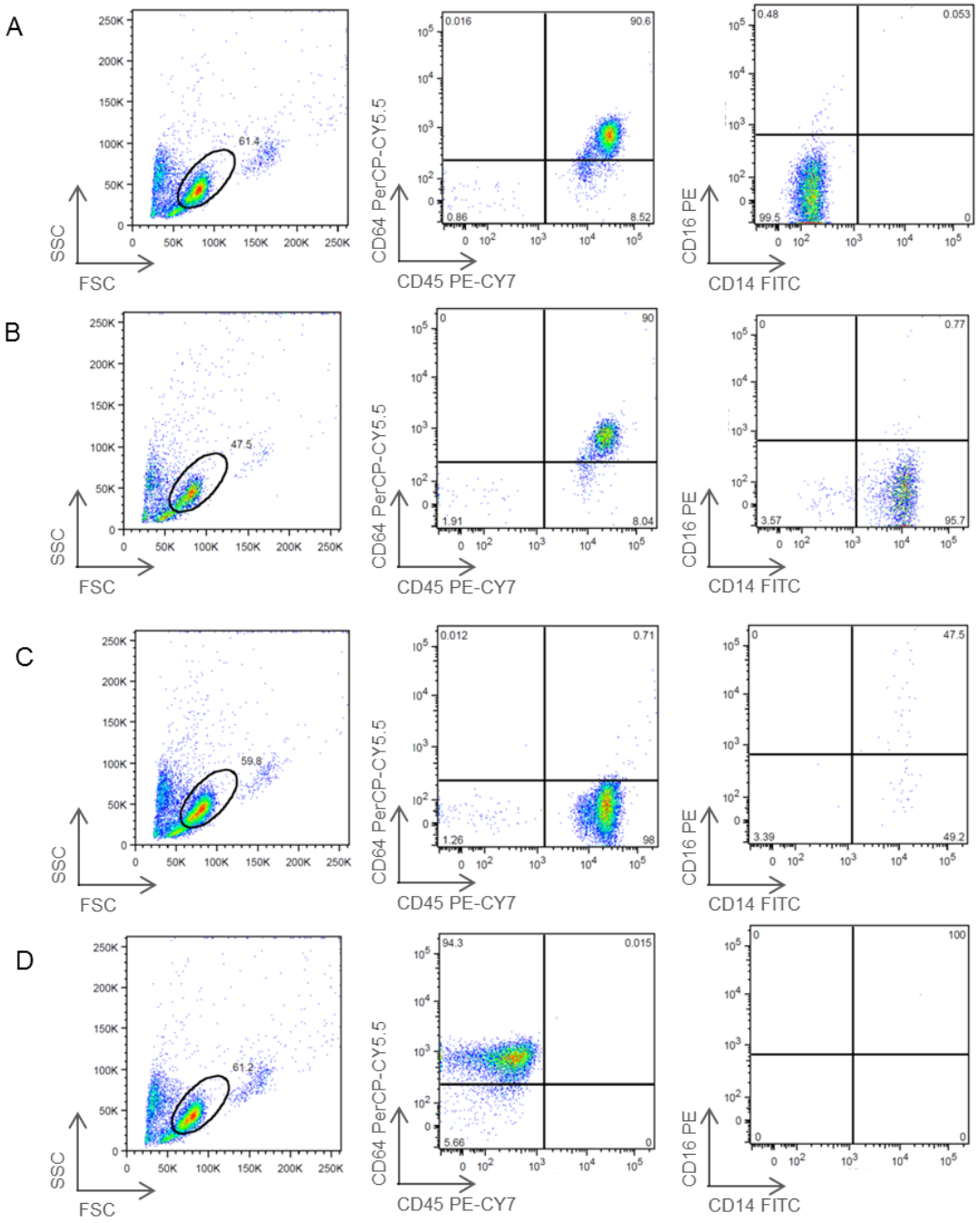
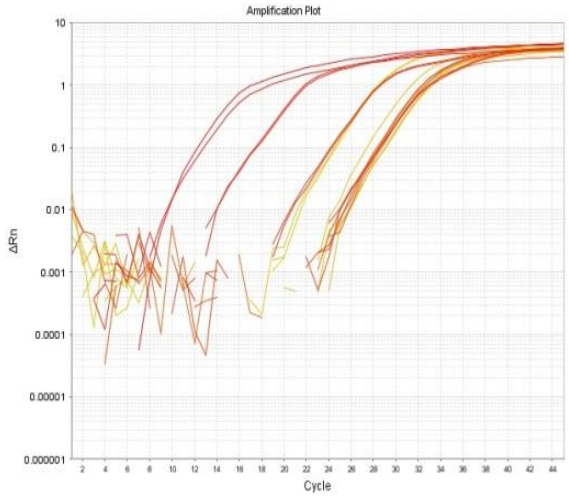


Figure 8.2. Fluorescence-minus-one controls for FACS analyses. FMO controls were used to identify and gate cells due to the multiple antibodies staining panel. The controls were set up by omitting the fluorochrome being analysed, FITC (A), PE (B), PerCP-CY5.5 (C) and PE-CY7 (D).

8.3 Appendix III - Results

A



B

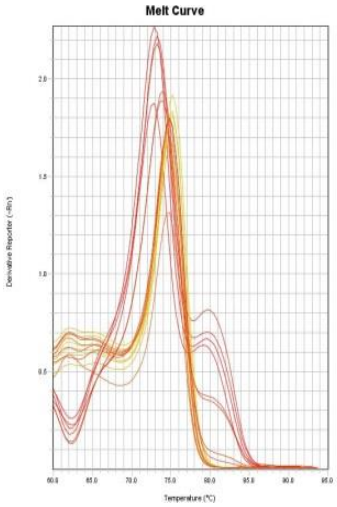


Figure 8.3. qPCR optimisation step with DMSO. One of the optimisation steps was to use DMSO as it lowers the T_m of the primer-template hybridization reaction and reduce secondary structures. However, this did not remove the primer dimers as shown in the amplification plot (A) and melting curve analysis (B).

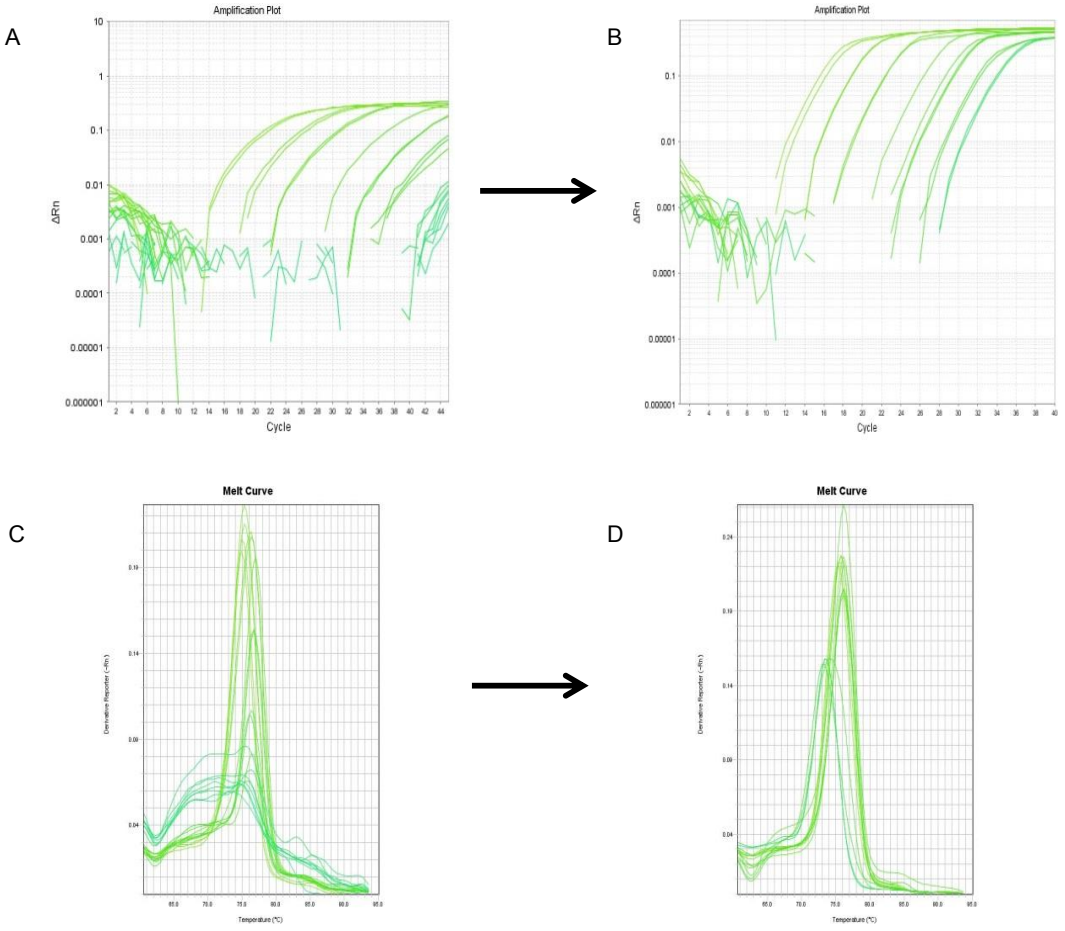


Figure 8.4. qPCR optimisation step with new, shorter primer sets. Shorter primers were designed as the original primer lengths were thought to have self-hybridized to produce primer dimers (A and C). The amplification curves were slightly better, however, primer dimers still appeared in the reactions (B and D).

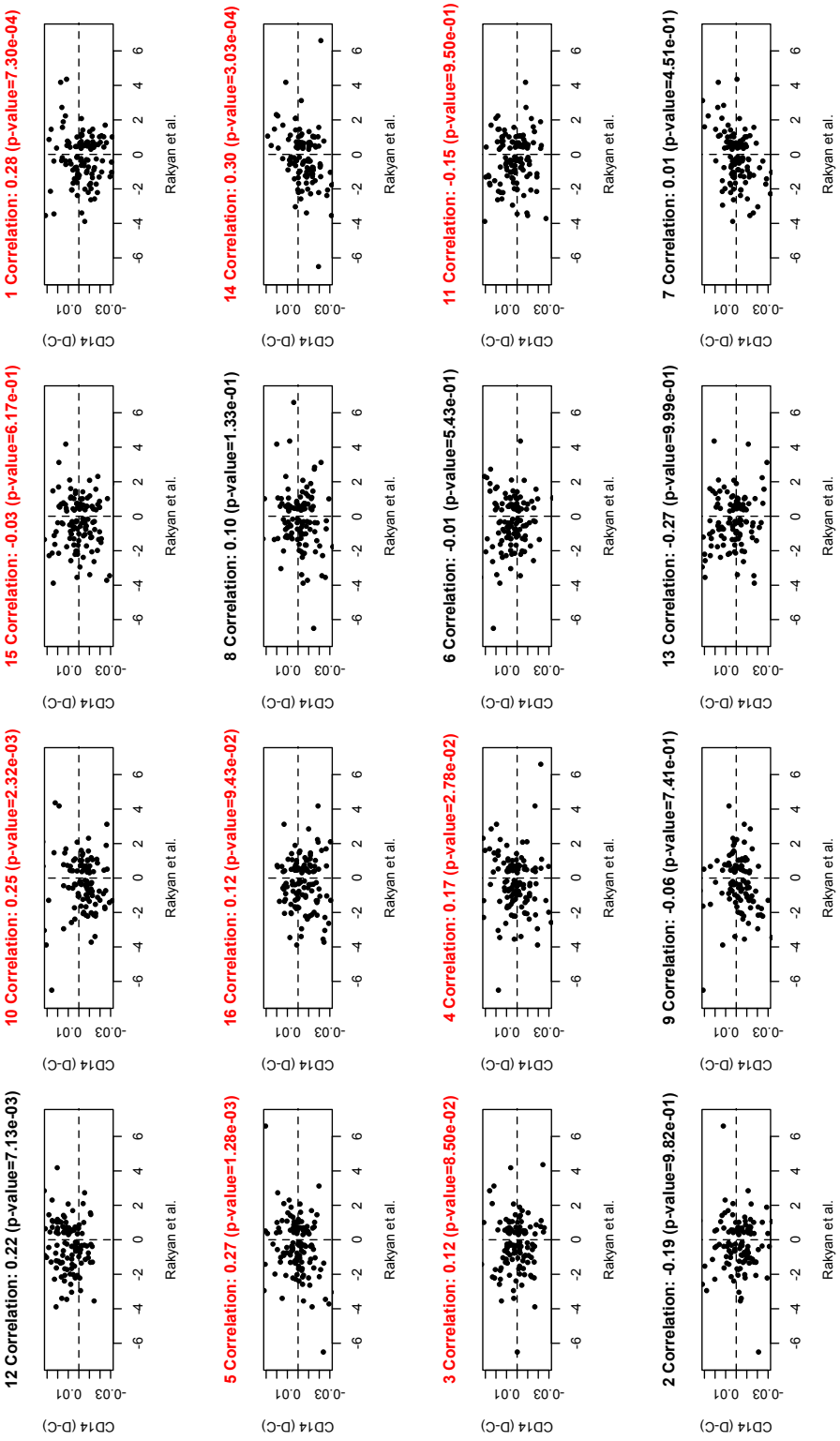


Figure 8.5. Individual correlation plots for each MZ twin pair from the Illumina450K analysis. The plots were ordered by age at diagnosis (left to right, top to bottom) with nine pairs of twins previously profiled using Illumina27K (red correlation).

Table 8.2. MVP calls for $CD4^+$ cells.

Probes	p.val	CHR	Arm	Gene	Feature		Enhancer
cg19845159	5.96E-07	8	q	ZNF34	5'UTR	Island	NA
cg20066488	1.19E-06	14	q	PPP1R13B	TSS1500	S_Shore	NA
cg08216099	1.67E-06	2	p	PXDN	Body		TRUE
cg21240640	3.93E-06	2	p	C2orf16	5'UTR		NA
cg09569760	3.93E-06	5	q	RNF130	Body		TRUE
cg00895997	3.93E-06	7	q	SLC13A1	TSS200		NA
cg25851842	3.93E-06	1	q	TGFB2	TSS200	Island	NA
cg05090359	3.93E-06	2	p	TPO	Body	N_Shore	NA
cg16297569	5.13E-06	1	p	GFI1	TSS1500	Island	NA
cg01587049	5.13E-06	14	q	MIR329-2	TSS1500		NA
cg25567280	6.56E-06	3	q	RNF13	Body		NA
cg27362718	8.34E-06	16	q	ZFPM1	Body		TRUE
cg25892296	1.05E-05	4	q	C4orf51	Body		TRUE
cg00303919	1.31E-05	1	p	REG4	3'UTR		TRUE
cg26894854	1.31E-05	1	q	SLC45A3	Body	N_Shore	TRUE
cg16492417	1.63E-05	14	q	FLRT2	5'UTR		NA
cg12828294	1.63E-05	8	q	HAS2	5'UTR		TRUE
cg21266698	1.63E-05	1	q	KIAA0040	5'UTR		NA
cg00333843	1.63E-05	16	q	MAF	3'UTR	S_Shelf	NA
cg08427977	2.01E-05	10	q	C10orf72	TSS1500	S_Shore	NA
cg10846775	2.01E-05	12	q	FAM101A	5'UTR	S_Shelf	NA
cg07955474	2.01E-05	16	q	IRF8	5'UTR	S_Shelf	NA
cg27130665	2.01E-05	11	p	LRRC4C	TSS1500		NA
cg00647232	2.01E-05	1	q	SEC16B	5'UTR		NA
cg08320359	2.47E-05	16	p	C16orf45	TSS200		TRUE
cg15030712	2.47E-05	7	p	CHN2	Body		TRUE
cg22821560	2.47E-05	1	q	KCNK2	TSS200		NA
cg00537210	2.47E-05	2	p	KLHL29	Body	S_Shore	TRUE
cg08827674	2.47E-05	17	q	LLGL2	5'UTR	Island	NA
cg00442646	2.47E-05	6	q	OSTCL	Body		NA
cg10084554	3.02E-05	10	q	ADAM12	Body		TRUE
cg22307508	3.02E-05	17	p	ASPA	TSS1500	S_Shelf	NA
cg26073101	3.02E-05	8	q	EEF1D	Body		NA
cg18703935	3.02E-05	6	p	HLA-H	Body	Island	NA
cg11287400	3.02E-05	3	p	MITF	Body		TRUE
cg23898305	3.02E-05	2	q	MYO7B	Body	N_Shelf	NA
cg21626573	3.02E-05	12	q	NCOR2	Body	N_Shore	TRUE
cg22929506	3.02E-05	2	q	PNKD	3'UTR	S_Shelf	NA
cg05352321	3.02E-05	10	q	SH3PXD2A	Body		TRUE
cg10734448	3.02E-05	6	p	TNXB	Body	N_Shelf	NA
cg10339152	3.66E-05	20	p	C20orf54	TSS1500		NA
cg03625136	3.66E-05	3	p	CACNA2D2	Body	S_Shore	NA
cg15330117	3.66E-05	10	p	FLJ45983	TSS1500	Island	NA
cg05346902	3.66E-05	19	q	MEIS3	Body	Island	TRUE
cg17895496	3.66E-05	15	q	PLA2G4F	TSS1500		NA
cg25290617	3.66E-05	17	q	RPTOR	3'UTR	Island	NA
cg17803589	3.66E-05	21	q	SLC19A1	Body	N_Shelf	NA
cg22842189	3.66E-05	3	p	THRB	5'UTR		TRUE
cg21535670	3.66E-05	8	q	TRAPPC9	Body		NA
cg07718813	3.66E-05	1	p	ZBTB48	Body	S_Shore	NA

Table 8.3. MVP calls for $CD14^{+}CD16^{-}$ cells.

Probes	p.val	CHR	Arm	Gene	Feature		Enhancer
cg10350689	1.63E-05	4	q	CXCL1	Body	S_Shore	NA
cg25409040	3.02E-05	11	p	CSTF3	TSS1500	S_Shore	NA
cg00955686	3.02E-05	1	p	MAP7D1	TSS1500	N_Shore	NA
cg03454705	4.42E-05	17	q	FOXK2	Body	S_Shore	NA
cg04557883	6.39E-05	6	p	C6orf145	TSS1500	S_Shore	NA
cg00031162	7.63E-05	17	p	TNFSF12	Body	S_Shore	TRUE
cg15472754	9.08E-05	3	q	MUC4	TSS1500		NA
cg17194243	0.000127	13	q	FAM48A	5'UTR	Island	NA
cg02974499	0.000127	11	q	FAU	TSS200	Island	NA
cg13681781	0.000127	9	q	QSOX2	Body		NA
cg14723284	0.00015	21	q	C21orf70	Body	S_Shelf	TRUE
cg15808604	0.00015	15	q	LOC283731	Body	N_Shore	NA
cg18449187	0.00015	19	q	NLRP7	TSS1500		NA
cg10476558	0.000175	20	q	TPX2	TSS200	N_Shore	NA
cg12243582	0.000175	7	q	ZNF138	TSS1500	N_Shore	NA
cg09837298	0.000205	14	q	ITPK1	Body		TRUE
cg13088038	0.000205	2	q	LOC440839	Body		TRUE
cg17926869	0.000239	19	p	CCDC124	5'UTR	S_Shore	NA
cg25877019	0.000239	7	p	FOXK1	3'UTR	S_Shore	NA
cg17869960	0.000239	6	p	PHACTR1	5'UTR		NA
cg22828990	0.000239	14	q	SERPINA1	5'UTR		TRUE
cg17589866	0.000278	2	p	ALK	Body		TRUE
cg15630071	0.000278	3	p	TRIM71	Body	Island	NA
cg06058311	0.000322	8	q	EFR3A	TSS1500	N_Shore	NA
cg11807539	0.000322	12	q	GALNT9	Body	N_Shore	NA
cg16156617	0.000322	6	p	SLC17A4	Body		NA
cg20973958	0.000373	13	q	ATP11A	Body	N_Shelf	TRUE
cg02047809	0.000373	14	q	COCH	TSS1500	N_Shore	TRUE
cg07464125	0.000373	10	q	KCNMA1	Body		NA
cg03585888	0.000373	19	q	MYH14	Body	S_Shore	NA
cg04043892	0.000373	12	q	PXN	Body	S_Shore	NA
cg00364357	0.000373	15	q	SNX22	TSS1500	Island	NA
cg00285317	0.00043	2	q	MLPH	Body		TRUE
cg27039218	0.00043	10	p	NRP1	Body		TRUE
cg13063900	0.00043	12	p	PDE3A	Body		TRUE
cg18262201	0.00043	10	p	PFKFB3	Body	Island	NA
cg26129110	0.000494	5	p	ADCY2	Body		TRUE
cg04391205	0.000494	2	q	C2orf88	TSS1500	N_Shore	NA
cg17459497	0.000494	2	p	COLEC11	Body		NA
cg07597706	0.000494	17	q	CUEDC1	5'UTR	S_Shore	NA
cg06215782	0.000494	12	p	EFCAB4B	Body		NA
cg05786348	0.000494	2	q	GPD2	5'UTR	S_Shelf	NA
cg12879013	0.000494	4	q	KIAA1211	5'UTR		NA
cg18368669	0.000494	4	p	PCGF3	Body	Island	NA
cg12094903	0.000494	6	p	PSMB8	3'UTR	N_Shelf	NA
cg21582758	0.000494	5	q	RASGEF1C	Body	S_Shore	NA
cg12006544	0.000494	13	q	RNF17	Body		NA
cg15392029	0.000494	11	q	TMEM136	TSS1500	N_Shore	NA
cg20555564	0.000568	14	q	C14orf133	TSS1500	Island	NA
cg08224238	0.000568	16	q	CA5A	Body		TRUE

Table 8.4. MVP calls for CD19^{+} cells.

Probes	p.val	CHR	Arm	Gene	Feature		Enhancer
cg01934142	5.96E-07	4	p	RFC1	TSS200	Island	NA
cg19866406	2.98E-06	1	p	SLC6A17	Body		TRUE
cg06899836	8.34E-06	6	q	CD164	TSS1500	Island	NA
cg05392764	1.05E-05	1	q	FCRLB	Body	Island	NA
cg16161418	1.05E-05	6	p	KIAA1949	3'UTR	S_Shelf	NA
cg13031595	1.31E-05	20	q	ARFGEF2	3'UTR		NA
cg04481722	1.31E-05	22	q	DGCR6	TSS200	Island	NA
cg08136747	1.31E-05	19	p	PGPEP1	TSS200	Island	NA
cg06454848	1.31E-05	19	q	ZNF260	5'UTR	Island	NA
cg14088052	2.01E-05	7	p	CCM2	TSS1500	N_Shore	NA
cg13529064	2.01E-05	7	p	URGCP	Body		NA
cg27629776	2.47E-05	20	q	FAM83C	TSS200	S_Shore	NA
cg21310336	2.47E-05	6	p	ZNF311	TSS1500		NA
cg03801030	3.02E-05	11	p	APBB1	TSS1500	S_Shore	NA
cg24863569	3.02E-05	5	q	COL23A1	Body		NA
cg19057227	3.02E-05	2	p	TTC15	Body		NA
cg06145508	3.66E-05	10	q	ABLIM1	Body		NA
cg07510303	3.66E-05	6	q	ADAT2	3'UTR		TRUE
cg03639021	3.66E-05	5	q	GNB2L1	TSS200	Island	NA
cg27622515	3.66E-05	16	p	LMF1	Body	N_Shore	NA
cg05786278	3.66E-05	2	p	MYT1L	5'UTR		NA
cg16248277	4.42E-05	10	q	FGF8	TSS1500	Island	NA
cg20460101	5.33E-05	7	q	CDK6	5'UTR	Island	NA
cg16141228	5.33E-05	5	q	CLINT1	TSS200	S_Shore	NA
cg17832805	5.33E-05	6	q	HMGN3	TSS200	Island	NA
cg01021488	5.33E-05	15	q	IGF1R	Body	Island	NA
cg08571639	5.33E-05	2	q	LRRFIP1	Body		TRUE
cg14204738	5.33E-05	1	q	RORC	TSS1500		NA
cg12484411	5.33E-05	8	q	TRHR	TSS200		NA
cg26178529	6.39E-05	6	q	ATG5	TSS200	Island	NA
cg09008417	6.39E-05	3	p	CCDC13	TSS200	Island	TRUE
cg01282921	6.39E-05	10	q	DLG5	Body		TRUE
cg16435469	6.39E-05	2	q	LYPD6	5'UTR		NA
cg01525976	6.39E-05	12	q	PTPRR	Body		TRUE
cg09728637	6.39E-05	18	p	TYMS	Body	S_Shore	NA
cg07148743	7.63E-05	5	p	BRD9	Body	N_Shore	NA
cg09301086	7.63E-05	2	p	COLEC11	Body	Island	NA
cg20356147	7.63E-05	3	q	EEFSEC	Body		TRUE
cg04342999	7.63E-05	5	q	MATR3	TSS200	Island	NA
cg02248037	7.63E-05	6	p	MICB	5'UTR	Island	NA
cg26646397	7.63E-05	5	q	NDST1	Body	S_Shelf	TRUE
cg27389262	7.63E-05	1	q	OTUD7B	TSS200	Island	NA
cg13363640	7.63E-05	11	p	SNORA54	TSS200		NA
cg20594671	7.63E-05	2	p	SNTG2	Body		NA
cg19334452	7.63E-05	19	p	TLE2	Body	Island	TRUE
cg00320608	7.63E-05	2	p	WDR43	TSS1500	N_Shore	NA
cg10982358	9.08E-05	1	p	CDKN2C	TSS200	S_Shore	NA
cg13560576	9.08E-05	9	p	ERMP1	TSS1500	S_Shore	NA
cg02426093	9.08E-05	6	q	FBXO5	TSS1500	S_Shore	NA
cg01382281	9.08E-05	19	q	FIZ1	Body	Island	NA

Table 8.5. Pyrosequencing results from SP1. Twin pairs have been omitted from the analysis due to failed checks in the assay. Difference in methylation between the twin pairs were calculated as follows: Non-T1D minus T1D. Negative values are in red.

	CpG1			CpG2		
Twin pair	**T1D**	**Non-T1D**	**Difference %**	**T1D**	**Non-T1D**	**Difference %**
2	81.06	100	18.94	59.89	66.62	6.73
3	89.34	78.31	-11.03	62.96	60.61	-2.35
4	90.13	95.56	5.43	50.42	69.45	19.03
6	89.43	84.91	-4.52	58.15	48.36	-9.79
8	89.16	89.3	0.14	82.92	76.32	-6.6
9	64.86	73.1	8.24	59.73	61.8	2.07
10	90.22	85.72	-4.5	67.76	82.55	14.79
11	99.4	68.48	-30.92	72.37	67.71	-4.66
12	94.23	97.34	3.11	68.55	74.87	6.32
13	52.55	80.1	27.55	37.39	69.85	32.46
14	92.32	87.74	-4.58	71.79	77.06	5.27
15	81.58	84.56	2.98	63.88	74.66	10.78
16	76.29	87.21	10.92	82.12	58.4	-23.72
17	83.38	79.07	-4.31	58.97	57.74	-1.23
18	96.48	100	3.52	84.2	59.85	-24.35
20	79.33	84.89	5.56	69.28	75.12	5.84
21	89.32	87.77	-1.55	58.86	78.75	19.89
22	84.15	62	-22.15	55.58	59.89	4.31
24	65.36	47.8	-17.56	34.75	41.16	6.41
25	82.89	93.26	10.37	58.11	56.38	-1.73
26	37.22	73.02	35.8	34.62	52.41	17.79
31	44.27	88.86	44.59	50.98	57.95	6.97
32	87.73	88.98	1.25	71.96	62.25	-9.71
33	84.76	88.65	3.89	65.91	74.43	8.52

Table 8.6. Pyrosequencing results from SP2. Twin pairs have been omitted from the analysis due to failed checks in the assay. Difference in methylation between the twin pairs were calculated as follows: Non-T1D minus T1D. Negative values are in red.

	CpG1			CpG2		
Twin pair	**T1D**	**Non-T1D**	**Difference %**	**T1D**	**Non-T1D**	**Difference %**
1	75.02	71.18	-3.84	55.79	60.21	4.42
2	73.37	80.22	6.85	59.37	74.24	14.87
3	70.03	80.99	10.96	48.81	67.48	18.67
4	71.81	91.93	20.12	70.34	59.87	-10.47
6	78.02	68.95	-9.07	81.85	49.14	-32.71
8	74.04	85.93	11.89	53.84	65.38	11.54
10	85.97	77.26	-8.71	69.69	73.97	4.28
12	74.01	75.82	1.81	73.25	64.01	-9.24
13	76.73	84.39	7.66	39.26	57.92	18.66
14	83.15	84.64	1.49	65.3	78.21	12.91
15	78.58	88.04	9.46	62.72	78.3	15.58
16	75.02	78.82	3.8	75.55	59.32	-16.23
17	75.48	73.94	-1.54	60.19	63.13	2.94
19	33.46	76.98	43.52	52.88	76.1	23.22
20	78.96	78.05	-0.91	62.89	54.59	-8.3
21	73.22	85.22	12	52.6	68.92	16.32
24	67.56	33.27	-34.29	69.53	33.49	-36.04
25	87.8	76.44	-11.36	58.69	64.11	5.42
26	31.91	53.33	21.42	38	39.62	1.62
30	80.44	73.83	-6.61	49.01	100	50.99
31	79.27	80.99	1.72	36.01	68.28	32.27
32	82	73.34	-8.66	62.82	61.01	-1.81
33	83.53	80.36	-3.17	63.07	63.96	0.89