Genome-wide analysis of the DNA methylation landscape in human tissues

Volkov, Petr

2016

Document Version:
Publisher's PDF, also known as Version of record

Genome-wide analysis of the DNA methylation landscape in human tissues

Petr Volkov

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Aula Auditorium, on September 14th, at 13:15.

Faculty opponent
Assistant Professor David Gomez-Cabrero,

Karolinska Institute,

Sweden

Organization LUND UNIVERSITY	Document name Doctoral Dissertation	
	Date of issue September 13th, 2016	
Author(s) Petr Volkov	Sponsoring organization	

Title and subtitle Genome-wide analysis of the DNA methylation landscape in human tissues

Abstract

The prevalence of type 2 diabetes and obesity is rapidly increasing worldwide. Both pancreatic islets and adipose tissue play an important role in metabolic processes in humans, and dysfunction in these tissues may contribute to the pathogenesis of type 2 diabetes and obesity. Epigenetics is thought to be one of the processes by which the environment interacts with cellular phenotypes. However, current knowledge about the role of epigenetics in type 2 diabetes and obesity remains limited. In 4 studies included in this thesis, we have shown that the DNA methylome in pancreatic islets and adipose tissue reflects environmental factors, metabolic phenotypes and type 2 diabetes status.

In study 1 we investigated whether the DNA methylome in human adipose tissue exhibit changes in DNA methylation in response to 6 months exercise intervention. Overall, we identified 17,975 CpG sites significantly altered before versus after the intervention. These alterations involve some of the genes previously associated with type 2 diabetes and obesity, including *TCF7L2* and *KCNQ1*.

Next, in study 2 we identified a set of CpG sites which are associated with age, BMI or HbA1c in human adipose tissue. CpG sites associated with age in adipose tissue were annotated to previously known blood-based biomarkers for age, such as *ELOVL2*. Moreover, we identified multiple genes, including such obesity loci as *FTO*, *ITIH5* and *CCL18*, where both DNA methylation and mRNA expression significantly correlated with BMI.

In study 3, we investigated the influence of genetic variation on DNA methylation in human adipose tissue and charted a map of SNP-CpG interactions in human adipose tissue. We identified interactions between some of the SNPs previously implicated in obesity and lipid biology and the DNA methylome in human adipose tissue. We also demonstrated that DNA methylation can play a mediating role between genetic variation and transcriptional activity.

In study 4, we used whole-genome bisulfite sequencing to chart the DNA methylome of human pancreatic islets, and to identify contiguous genomic regions significantly different between donors with or without type 2 diabetes. We found that in human pancreatic islets differentially methylated regions are associated with transcriptional activity and are overrepresented in regulatory regions such as enhancers and transcription factor binding sites.

Overall, these studies suggest that DNA methylation plays a role in the way the environment affects cellular function in both human adipose tissue and pancreatic islets.

Key words DNA methylation, type 2 diabetes, WGBS, Illumina 450k, mQTL, genetics, adipose tissue, human pancreatic islets

Classification system and/or index terms (if any)

Supplementary bibliographical information		Language English
ISSN and key title 1652-8220		ISBN 978-91-7619-332-7
Recipient's notes	Number of pages	Price
	Security classification	

Signature  Date 2016-08-03

Genome-wide analysis of the DNA methylation landscape in human tissues

Petr Volkov

LUND
UNIVERSITY

Coverphoto by Linn Gillberg

Faculty of Medicine
Department of Clinical Sciences

ISBN 978-91-7619-332-7
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2016

To my grandparents

Content

List of publications

Publications included in the thesis

1. Rönn T, **Volkov P**, Davegårdh C, Dayeh T, Hall E, Olsson AH, Nilsson E, Tornberg Å, Dekker N, Eriksson K-F, Jones HA, Groop L, Ling C. *A Six Months Exercise Intervention Influences the Genome-wide DNA Methylation Pattern in Human Adipose Tissue*. PLoS Genetics. 2013;9(6).
2. **Volkov P**, Rönn T, Gillberg L, Kokosar M, Perfilyev A, Jacobsen AL, Jørgensen SW, Brøns C, Jansson P-A, Eriksson K-F, Pedersen O, Hansen T, Groop L, Stener-Victorin E, Vaag A, Nilsson E, Ling C. *Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood*. Hum Mol Genet. 2015 Jul 1;24(13):3792–813.
3. **Volkov P**, Olsson AH, Gillberg L, Jørgensen SW, Brøns C, Eriksson K-F, Groop L, Jansson P, Nilsson E, Rönn T, Vaag A, Ling C. *A genome-wide mQTL analysis in human adipose tissue identifies genetic variants associated with DNA methylation, gene expression and metabolic traits*. PLOS ONE. 2016;11: e0157776. doi:10.1371/journal.pone.0157776
4. **Volkov P**, Bacos K, Ofori JK, Esguerra JLS, Eliasson L, Rönn T, Ling C. Whole-genome bisulfite sequencing of human pancreatic islets reveals novel differentially methylated regions in type 2 diabetes pathogenesis. Submitted

Publications not included in the thesis

1. Yang BT, Dayeh TA, **Volkov P**, Kirkpatrick CL, Malmgren S, Jing X, Renström E, Wollheim CB, Nitert MD, Ling C. *Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes*. Molecular endocrinology (Baltimore, Md). 2012 Jul;26(7):1203–1212.
2. Nitert MD, Dayeh T, **Volkov P**, Elgzyri T, Hall E, Nilsson E, Yang BT, Lang S, Parikh H, Wessman Y, Weishaupt H, Attema J, Abels M, Wierup N, Almgren P, Jansson P-A, Rönn T, Hansson O, Eriksson K-F, Groop L, Ling C. *Impact of an Exercise Intervention on DNA Methylation in Skeletal Muscle From First-Degree Relatives of Patients With Type 2 Diabetes*. Diabetes. 2012 Dec 1;61(12):3322–32.
3. Dayeh TA, Olsson AH, **Volkov P**, Almgren P, Rönn T, Ling C. Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia. 2013 May;56(5):1036–1046.
4. Rönn T, **Volkov P**, Tornberg A, Elgzyri T, Hansson O, Eriksson K-F, Groop L, Ling C. Extensive changes in the transcriptional profile of human adipose tissue including genes involved in oxidative phosphorylation after a 6-month exercise intervention. Acta Physiologica. 2014;211(1):188–200.
5. Dayeh T, **Volkov P**, Salö S, Hall E, Nilsson E, Olsson AH, Kirkpatrick CL, Wollheim CB, Eliasson L, Rönn T, Bacos K, Ling C. Genome-Wide DNA Methylation Analysis of Human Pancreatic Islets from Type 2 Diabetic and Non-Diabetic Donors Identifies Candidate Genes That Influence Insulin Secretion. PLoS Genetics. 2014;10(3).
6. Nilsson E, Jansson PA, Perfilyev A, **Volkov P**, Pedersen M, Svensson MK, Poulsen P, Ribel-Madsen R, Pedersen NL, Almgren P, Fadista J, Rönn T, Pedersen BK, Scheele C, Vaag A, Ling C. *Altered DNA Methylation and Differential Expression of Genes Influencing Metabolism and Inflammation in Adipose Tissue From Subjects With Type 2 Diabetes*. Diabetes. 2014 Sep 1;63(9):2962–76.
7. Olsson AH, **Volkov P**, Bacos K, Dayeh T, Hall E, Nilsson EA, Ladenvall C, Rönn T, Ling C. Genome-Wide Associations between Genetic and Epigenetic Variation Influence mRNA Expression and Insulin Secretion in Human Pancreatic Islets. PLoS Genetics. 2014;10(11).
8. Hall E, **Volkov P**, Dayeh T, Esguerra JL, Salö S, Eliasson L, Rönn T, Bacos K, Ling C. Sex differences in the genome-wide DNA methylation

pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. Genome biology. 2014;15(12):522.

9. Hall E, **Volkov P**, Dayeh T, Bacos K, Rönn T, Nitert MD, Ling C. Effects of palmitate on genome-wide mRNA expression and DNA methylation patterns in human pancreatic islets. BMC Medicine. 2014;12(1).

10. Rönn T, **Volkov P**, Gillberg L, Kokosar M, Perfilyev A, Jacobsen AL, Jørgensen SW, Brøns C, Jansson P-A, Eriksson K-F, Pedersen O, Hansen T, Groop L, Stener-Victorin E, Vaag A, Nilsson E, Ling C. *Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood*. Hum Mol Genet. 2015 Jul 1;24(13):3792–813.

11. Agardh E, Lundstig A, Perfilyev A, **Volkov P**, Freiburghaus T, Lindholm E, Rönn T, Agardh C-D, Ling C. Genome-wide analysis of DNA methylation in subjects with type 1 diabetes identifies epigenetic modifications associated with proliferative diabetic retinopathy. BMC Medicine. 2015;13(1).

12. Bacos K, Gillberg L, **Volkov P**, Olsson AH, Hansen T, Pedersen O, Gjesing AP, Eiberg H, Tuomi T, Almgren P, Groop L, Eliasson L, Vaag A, Dayeh T, Ling C. *Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes*. Nature Communications. 2016;7.

13. Gillberg L, Perfilyev A, Brøns C, Thomasen M, Grunnet LG, **Volkov P**, Rosqvist F, Iggman D, Dahlman I, Risérus U, Rönn T, Nilsson E, Vaag A, Ling C. *Adipose tissue transcriptomics and epigenomics in low birthweight men and controls: role of high-fat overfeeding*. Diabetologia. 2016;59(4):799–812.

Abbreviations

BMI	body mass index
CIT	causal inference test
CGI	CpG island
CpG	cytosine guanine dinucleotide
DMRs	differentially methylated regions
DNMT	DNA methyltransferase
eQTL	expression quantitative trait loci
EWAS	epigenome-wide association study
FDR	false discovery rate
FFA	free fatty acids
GWAS	genome-wide association studies
HbA1c	haemoglobin A1c
HOMA-B	homeostatic model assessment of beta-cell function
HOMA-IR	homeostatic model assessment of insulin resistance
LD	linkage disequilibrium
MODY	maturity onset diabetes of the young
mQTL	methylation quantitative trait loci
RMA	robust multichip average
RRBS	reduced representation bisulfite sequencing
SNPs	single nucleotide polymorphisms
TPM	transcript per million
TES	transcription end site
TSS	transcription start site
UTR	untranslated region
WGBS	whole-genome sequencing of bisulfite treated DNA
WHO	World Health Organisation
WHR	waist-hip ratio

Introduction

Heritable genetic markers are one of the sources of phenotypic variation of complex traits. Heritability is a measure which characterizes the genetic influence on phenotypic variation [1]. There are several methods to estimate heritability, some based on computational methods, others, most prominent so far, analyse difference in concordance rate between pairs of monozygotic and dizygotic twins [1,2]. According to heritability studies, certain degree of phenotypic variance of many important phenotypic traits cannot be fully attributed to genetics alone [1–3].

It is well known that environmental factors play an important role in the pathogenesis of complex diseases [4]. Lifestyle patterns underwent major changes in the past several decades, which adversely affected epidemiology of multiple important metabolic traits. For example, type 2 diabetes, having heritability estimates from 20% to 80% [5], is an exponentially increasing pandemic with number of cases expected to reach 624 million in year 2040 [6]. This drastic increase was attributed to a sedentary lifestyle and excessive caloric intake, highly prevalent in modern societies [7]. Obesity is another major example of such a disease with both environmental and genetic influences, and its prevalence doubled from 1980 to reach 600 million cases in 2014, affecting up to 50% in certain populations [8]. Clinical complications of type 2 diabetes, for example diabetic retinopathy or kidney or nerve damage, and obesity, such as stroke and heart disease, present an increasing burden on healthcare systems [9].

Type 2 diabetes and obesity

Diabetes mellitus is a disease which is commonly characterized by beta-cell dysfunction, insulin resistance and increased blood glucose. According to World Health Organisation (WHO) standards, diabetes is diagnosed by either having fasting glucose levels > 7 mmol/L, 2-h plasma glucose after a 75-g oral glucose tolerance test of >11.1 mmol/L or having haemoglobin A1C (HbA1c) > 6.5% [10]. Diabetes mellitus is divided on multiple subtypes, including, but not limited to, type 1 and type 2 diabetes, gestational diabetes, which is developed by women during pregnancy, and several forms of monogenic diabetes such as maturity onset diabetes of the young (MODY) [10].

Insulin is a hormone required for maintenance of normal blood glucose level (normoglycaemia). It is secreted from beta-cells in response to elevated glucose levels after food intake, and functions as a signal to its target tissues such as liver, adipose tissue and skeletal muscle to initiate glucose uptake [11]. The insulin-secreting beta-cells are part of the pancreatic islets of Langerhans along with several other types of endocrine cells, including alpha-cells, delta-cells, PP-cells and epsilon-cells [12]. Alpha-cells secrete glucagon, which regulates normoglycaemia by stimulating the conversion of liver glycogen into glucose. Delta-cells are responsible for production of somatostatin, which is an inhibitor that coordinates glycogen and insulin secretion [13]. PP-cells produce pancreatic-polypeptide, which, among other functions, also inhibits insulin secretion [14]. Finally, epsilon-cells secrete ghrelin, a hormone which regulates insulin secretion as well as food intake initiation [15].

Type 2 diabetes is the most prevalent form of diabetes mellitus [10]. It is characterized by insulin resistance, i.e. inability of cells of target tissues to respond to insulin, which results in decreased glucose uptake by skeletal muscle and adipose tissue and increased glucose output from the liver, and thereby hyperglycaemia [10]. Initially, pancreatic beta-cells are compensating by insulin over-production and secretion [16]. However, overproduction of insulin by beta-cells is one of the causes which eventually lead to reduction in beta-cell mass [17], which can also occur due to high glucose concentrations [18]. Eventually, when the beta-cells are no longer able to compensate with elevated insulin secretion, diabetes manifests.

High glucose levels may also damage the liver, kidneys, and in due course lead to microvascular complications such as diabetic retinopathy or neuropathy [19]. Treatments for type 2 diabetes include, among many others, metformin, commonly used to increase insulin sensitivity and preserve beta-cell function, and insulin injections in cases when beta-cells become dysfunctional [20].

Obesity is characterized by excess in fat mass, and is most commonly diagnosed by having a body mass index (BMI) >= 30 kg / m^2 [21]. Dysfunction of adipose tissue is linked to both obesity and type 2 diabetes [22] and obesity is known to be associated with insulin resistance [23], making adipose tissue a key tissue for understanding type 2 diabetes pathogenesis [24].

Adipose tissue is an organ responsible for long-term energy storage as well as metabolic homeostasis [25]. It is composed by multiple cell types, including adipocytes, which are cells responsible for energy storage in form of triglycerides [26,27]. Adipose tissue contributes to glucose and free fatty acids (FFA) metabolism by secreting several proteins, including, but not limited to, leptin and adiponectin hormones [27]. Like ghrelin, leptin is involved in food intake initiation and hunger control, by interaction with hypothalamus [28]. It also

contributes to regulation of glucose levels and insulin resistance [29–31]. Reduced leptin secretion is associated with obesity. Adipocytes can experience hypertrophy in response to metabolic stress, such as long-term overfeeding [32], which impairs their function, including reduced leptin secretion [33]. Moreover, resistance of hypothalamus to leptin signalling has been linked to obesity [34].

Genetics of type 2 diabetes and obesity

Advances in mapping human genetic variation, such as the HapMap [35] or 1000 Genomes projects [36], and the development of technologies such as DNA microarrays, made genome-wide screening of common markers in large cohorts possible. Genome-wide association studies (GWAS) contributed greatly to understanding genetic components of many complex traits. GWAS are usually performed by analysing statistical associations between directly genotyped and imputed single nucleotide polymorphisms (SNPs) with a phenotypic trait of interest and selecting associations with p-values below genome-wide significance threshold of $p < 5\times10^{-8}$ [37,38].

A family history of type 2 diabetes greatly increases the risk of developing the disease later in life – by 70% for individuals with both parents being affected by the disease and by 40% for individuals with one type 2 diabetic parent [39]. Linkage studies in 1990s and early 2000s identified several loci associated with type 2 diabetes, such as *CAPN1* and *TCF7L2* genes [40,41]. A breakthrough in understanding genetics of type 2 diabetes came in 2007, when several type 2 diabetes GWAS uncovered a genome-wide significant SNP close to *TCF7L2*, replicating linkage-based discoveries [5]. Since then, 153 variants, annotated to more than 120 genes, have been found to be associated with type 2 diabetes [42]. Variants with highest effect sizes are located in or close to *TCF7L2* (odds-ratio (OR) for risk-allele carriers 1.4), *KCNQ1* (OR 1.4) and *CDKN2A/B* (OR 1.3) [43].

GWAS were performed also for proxy measures of obesity such as body mass index (BMI) and waist-hip ratio (WHR), and identified 97 genome-wide significant loci associated with BMI and 49 with WHR [44,45]. Strongest genome-wide associations were found for variants in the *FTO* gene, with OR as high as 1.67 [44]. SNPs in or close to *FTO* were also identified in type 2 diabetes GWAS [46].

Standard practice in GWAS is to link identified genetic markers to genes based on genomic distance [47]. These markers can often be found to be located in the intronic or intergenic regions and, therefore, often do not modify the coding sequence of the gene [48]. Identifying causal variants based on statistical associations, as well as understanding underlying mechanisms of their action, often requires a combination of molecular and computational approaches. For

example, a study which identified a causal variant in the *FTO* gene analysed cross-species conservation patterns of transcription factor binding sites (TFBS) as well as utilized genome editing techniques [49]. This study, among others, points to importance of considering genetic markers in its surrounding context of TFBS and regulatory regions.

Both type 2 diabetes and obesity are complex diseases, meaning phenotypic variance is thought to be due to a combination of multiple genetic and environmental factors as well as interactions between them. Cellular phenotype is partly defined by gene expression, which is in turn coordinated by regulatory proteins including transcription factors [50]. However, transcription factors alone cannot address mechanisms by which the environment affects cellular circuitry [51]. Epigenetics is commonly thought to be a mechanism that can play a proxy role between environmental exposure and cellular phenotypes [52,53].

Epigenetics and DNA methylation

Epigenetics is an umbrella term for several biochemical phenomena which are broadly defined as modifications of a DNA molecule that cause changes of DNA function without involving alterations of the genomic sequence [54]. Several known biochemical processes are classified as epigenetics. One involves more than 100 different modifications of histone proteins such as, for example, methylation of lysine 4 on histone 3 (H3K4me and H3K4me3), which have been shown to be associated with chromatin structure and regulatory regions of the genome [55]. DNA methylation is another commonly studied epigenetic phenomenon.

DNA methylation is a process of addition of a methyl group to a cytosine base of a DNA molecule [56]. In differentiated mammalian cells, DNA methylation mostly occurs at cytosine-guanine dinucleotides, so-called CpG sites, while in plants non-CpG DNA methylation is also common [56]. The human genome contains approximately 28 million CpG sites [57] and, normally, between 70% to 80% of the CpG sites in the human genome are methylated [58]. Unmethylated CpG sites tend to occur in functional genomic regions such as promoters or enhancers [58,59]. DNA methyltransferase-3A and 3B (DNMT3A and 3B) are generally responsible for establishing *de novo* DNA methylation, while DNA methylation maintenance mainly seems to occur due to DNA methyltransferase-1 (DNMT1) [60]. Removal of DNA methylation marks can occur either passively, when DNA methylation is not preserved during DNA replication, for example due to a dysfunction of DNA methylation mechanisms, or actively, through conversion of 5-methylcytosines to 5-hydroxymethylcytosines by enzymes from the ten-eleven

translocation (TET) family [61] and subsequent conversion of 5-hydroxymethylcytosines to cytosines [62].

The relationship between tissue-specific DNA methylation and gene expression is currently being a subject of a scientific interrogation. It is well known that around 70% of all human gene promoters contain ~1kb long clusters of CpG sites, so-called CpG islands (CGIs) [58], and DNA methylation in CGI promoters is associated with transcriptional repression [58]. However, excluding cases such as imprinted genes and genes turned off by X-chromosome inactivation in females during development [63], DNA methylation in CGI promoters rarely occurs *de-novo,* due to enrichment of nucleosome depleted regions (signified by H3K4me3 histone mark) [58,64] (**Figure 1**). In the meantime, DNA methylation levels in non-CGI promoters seems more dynamic [65].

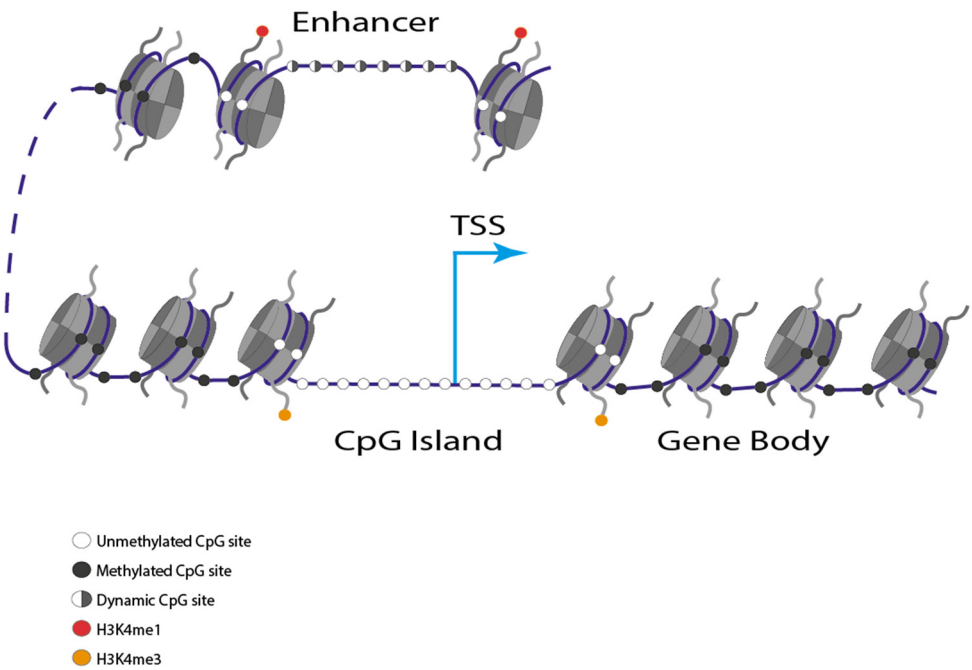


Figure 1. Genomic distribution of CpG methylation
Overall genomic landscape of DNA methylation. Most of the genome is methylated, enhancers are more dynamic in DNA methylation levels, CpG islands are commonly unmethylated and gene bodies often exhibit high levels of DNA methylation. Histone modification H3K4me3 is marking a nucleosome depleted region, which is associated with low activity of DNMTs.

DNA methylation was previously thought to be a solely repressive mark, and altered DNA methylation of promoter regions was previously mainly linked to diseases such as cancer [66]. However, expressed genes tend to exhibit higher

DNA methylation levels in their gene bodies [67], including higher DNA methylation in exons compared to introns and sharp changes in methylation levels in exon-intron boundaries [68], which suggests that DNA methylation can play a role in splicing mechanisms or alternative splicing events (**Figure 1**). Depending on the context, DNA methylation can either actively affect gene expression or be consequential to other factors [69]. All facts together point to the importance of considering DNA methylation marks in a surrounding genomic context rather than independently, and knowledge of the epigenetic landscapes of multiple tissues may help to unravel the mechanisms by which non-coding variants affect cellular phenotypes [49].

DNA methylation and type 2 diabetes

DNA methylation has previously been found to be potentially involved in both type 2 diabetes and obesity. For example, a study of *PPARGC1A,* a gene involved in glucose metabolism by regulating mitochondrial function, identified that it is differentially expressed in human pancreatic islets between diabetic and non-diabetic donors, and at the same time is potentially repressed by increased DNA methylation [70]. An epigenome-wide association study (EWAS) of human pancreatic islets analysed differential DNA methylation between type 2 diabetic and control donors and identified CpG sites in genes previously linked to type 2 diabetes, including *TCF7L2*, *FTO* and *KCNQ1* [71]. In an EWAS in blood of 5465 individuals, 37 CpG sites were identified to be associated with BMI, covering genes such as *CPT1A*, *ABCG1*, and *SREBF1* [72]. Additionally, differential DNA methylation patterns have also been found in the liver, adipose tissue and skeletal muscle from subjects with type 2 diabetes compared with non-diabetic controls [24,73,74]. Other studies have further hypothesized that DNA methylation can be one of the mechanisms of how the intrauterine environment affects cellular phenotypes and risk for different diseases later in life [75].

Epigenetic markers are known to be associated, and potentially affected, by genetic factors in both *cis* and *trans* [76] . SNPs statistically associated with histone modifications have been shown to localize in genomic regulatory regions and transcription factor binding sites [77], and even to switch these sites from an active to a repressed chromatin state [78]. These SNPs were also hypothesized to affect the surrounding chromatin state and to be a potential mechanism for expression quantitative trait loci (eQTLs) of non-coding sequences [76].

Moreover, multiple so-called methylation quantitative trait loci (mQTL) studies in various tissues, such as a study by Hannon et al. in brain [79] and Olsson et al. in human pancreatic islets [80], established maps of genetic-epigenetic interactions in both *cis* and *trans*. When treated as quantitative traits, many CpG sites were shown

to have high (up to 0.95) heritability estimates and, therefore, to be potentially under genetic control [81]. Similar to histone-QTL studies, significant mQTL loci are enriched in enhancer regions, as well as underrepresented in CpG island promoters [76]. However, since the epigenetic patterns are known to be tissue specific [59], the relationship between DNA methylation and genetic variation in adipose tissue, a key tissue for obesity and type 2 diabetes, remains an important area to study.

Bioinformatics of DNA methylation

Bisulfite conversion is a chemical reaction commonly used to differentiate between methylated and unmethylated cytosines in the DNA sequence [82]. During the treatment of the DNA with sodium bisulfite cytosine is converted to uracil, while 5-methylcytosine is not. Next, after polymerase chain reaction (PCR) uracil is converted to thymine. This fact is utilized for the identification of DNA methylation with the use of microarray and sequencing technologies, as thymine bases in bisulfite-converted DNA represent unmethylated cytosines in the DNA prior to bisulfite conversion, while cytosine residues point to methylated DNA.

Oligonucleotide hybridization of bisulfite converted DNA is one of the most commonly used technologies to investigate DNA methylation. There are several microarray solutions that target DNA methylation, such as the Illumina arrays (Illumina, San Diego, CA, USA) [83]. During our research we utilized the most comprehensive method available, Illumina Infinium HumanMethylation450 BeadChip (Illumina 450k array), which is a 12-sample microarray chip that incorporates 482,421 oligonucleotide probes designed to measure CpG methylation across the genome, as well as 3,091 non-CpG DNA methylations sites and 65 control SNPs [84]. CpG sites targeted by the chip cover 21,231 out of 21,474 UCSC RefSeq genes and 96% of UCSC CGIs [84].

DNA methylation probes utilized by the array are designed using two different assay chemistries [84]. First type, Infinium type I, covers 135,501 sites, and utilizes two bead types for methylated and unmethylated state of a target CpG site, respectively (**Figure 2**). Second type, Infinium type II, which covers the remaining 350,076 sites, has only one bead type corresponding to both DNA methylation states based on single base extension (**Figure 2**) [84].

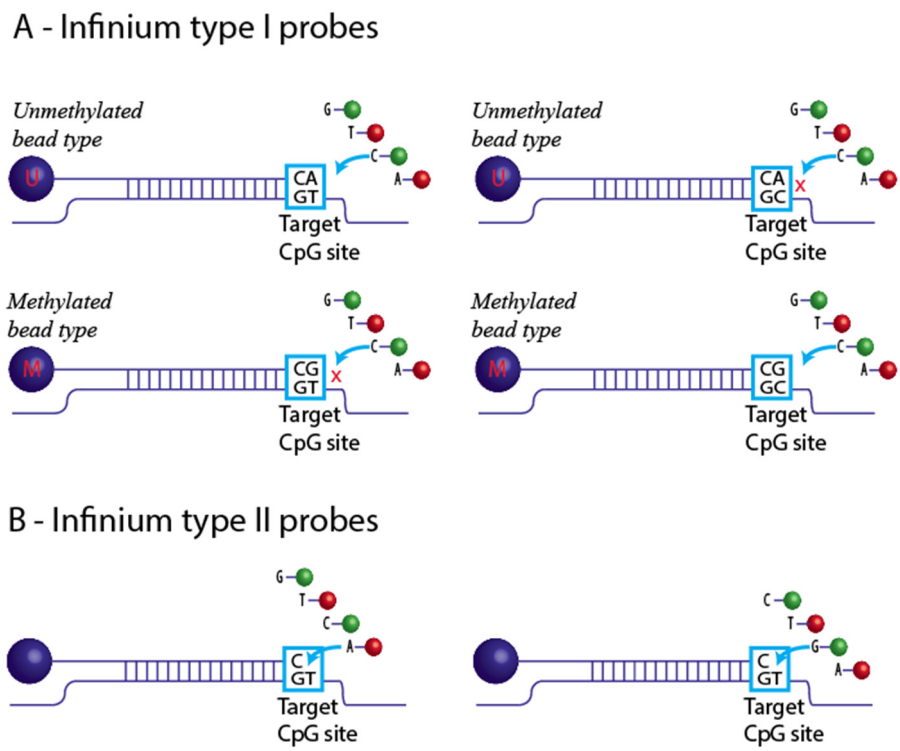


Figure 2. Design of Infinium type I and Infinium type II probe types

A – Infinium type I probe design incorporates 4 probes per CpG locus, 2 for methylated, and 2 for unmethylated state. Probes in methylated and unmethylated states contain different bead types, and CpG methylation is measured using a single colour channel.

B – Infinium type II probe design. Each CpG locus is measured by 2 probes in 2 different colours by using single-base extension.

Despite its widespread use, certain bioinformatic challenges are associated with the analysis of Illumina 450k array data.

DNA methylation is commonly represented as a proportion of cells methylated at a current genomic locus among the overall number of cells in a sampled material. This measure is also referred to as beta-value [85]. However, a study by Du et al. [85] reported that if DNA methylation, as measured by Illumina 450k array, is expressed in beta-values, then it exhibits heteroscedastic properties in hyper- and hypomethylated regions of the overall distribution. Therefore, additional transformation of beta-values to an alternative methylation measure was required.

Next, the difference in assay chemistry between type I and type II probes can be a source of a potential bias in the DNA methylation estimates. Indeed, it has been shown that the distribution of DNA methylation values differs between probe types (**Figure 3**) [86].

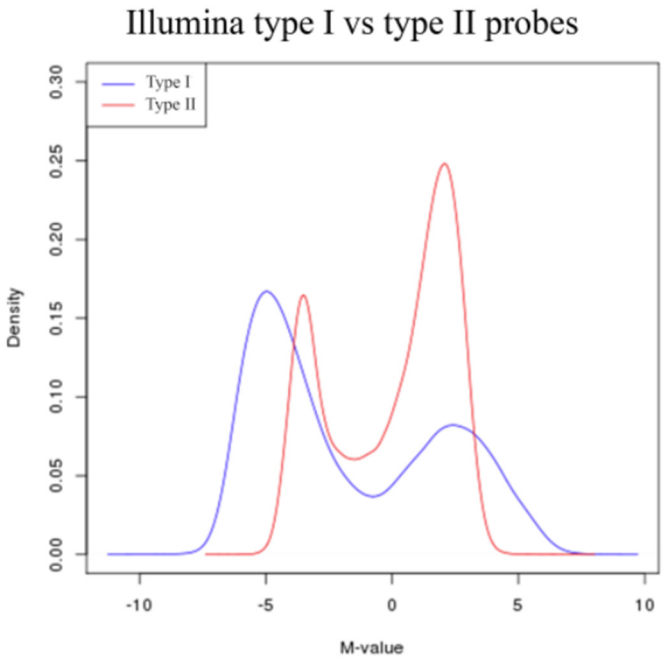


Figure 3. Distributions of DNA methylation values of Infinium type I and type II probes

This figure depicts distributions of DNA methylation measured by Infinium type I and Infinium type II probes separately. Here, Infinium type I probes are more sensitive to detect hypo- or hyper-methylation.

Another problem would be a potential cross-hybridization of probe sequences to multiple genomic locations as well as SNPs in probe targets. In fact, ~8.5% of Infinium Type I probes and ~5.1% of Infinium Type II probes are shown to potentially cross-hybridize to multiple genomic locations [87]. Moreover, ~14% of all probes target a CpG site with a known SNP [88].

Therefore, in order to perform biological investigations using Illumina 450k array in the course of this thesis, one of the requirements was to develop a bioinformatic pipeline to address aforementioned issues.

Despite the fact that Illumina 450k array allows "relatively quick" genome-wide DNA methylations screening, its coverage remains limited. First, only ~1.5% of 28 million CpG sites of the human genome are targeted by the array, which complicates a study of co-occurrence of DNA methylation in contiguous genomic regions, while annotation of contiguous regions of DNA methylation, and not of single CpG sites, is essential for understanding of how DNA methylation fit into surrounding genetic and regulatory context. For example, a recent study identified 27 type 2 diabetes risk loci in contiguous differentially methylated regions (DMRs) between diabetic and non-diabetic mice [89]. However, for many genes

only a few CpG sites are interrogated by the Illumina 450k array. Moreover, regulatory regions such as enhancers are also not fully covered by the array [90].

Therefore, to deepen the coverage of the analysis of the DNA methylation landscape, a more comprehensive approach to measure genome-wide DNA methylation patterns are needed.

There are several high coverage methods for genome-wide analysis of DNA methylation. For example, reduced representation bisulfite sequencing (RRBS) utilizes restriction enzyme digestion followed by bisulfite sequencing and allows interrogation of ~10% of genomic CpG sites [91]. Another method, MeDIP-seq, based on sequencing of immunoprecipitated DNA, not only can be used to measure DNA methylation of up to ~87% genomic CpG sites, but also allows for distinction between 5-methylcytosine and 5-hydroxymethylcytosine [92], which is another epigenetic modification that occurs on cytosine residues [93] and not distinguishable by bisulfite conversion. However, MeDIP-seq was shown to have lower accuracy than RRBS [92]. The most comprehensive method to date utilizes whole-genome shotgun sequencing of bisulfite treated DNA. Recent technological advances greatly reduced the costs of DNA sequencing. Whole-genome sequencing of bisulfite treated DNA (WGBS) potentially allows interrogation of every genomic CpG site. However, due to the fact that methylated sites are unevenly distributed in the genome, and that during bisulfite treatment unmethylated cytosines are converted to thymine, many contiguous genomic regions of the DNA after the reaction can potentially consist of only three residue types instead of four, which reduces sequence complexity. Therefore, sequence aligners of bisulfite converted reads should take into account various scenarios, as DNA methylation can be present or absent at a single genomic locus [94].

Most of the existing library preparation protocols for DNA sequencing, such as NEXTflex™ Bisulfite Library Prep Kit, produced by Bioo Scientific (Austin, TX, USA), require large amounts of input DNA, which is sometimes difficult to acquire from human tissues. Novel library preparation kits, such as EpiGnome™ Methyl-Seq kit (Illumina), require much smaller inputs. However, there is a necessity for technical validation of these technologies by means of previously established methods such as Illumina 450k array and pyrosequencing.

Aims

Study 1

To investigate human adipose tissue epigenome dynamicity in response to prolonged exercise exposure in subjects with or without a family history of type 2 diabetes.

Study 2

To identify DNA methylation marks that are associated with age, BMI and HbA1c in human adipose tissue.

Study 3

To investigate potential interactions between genetic variation and DNA methylation in human adipose tissue and identify whether those interactions jointly affect mRNA expression and metabolic phenotypes.

Study 4

To characterize the whole-genome DNA methylation landscape in human pancreatic islets and identify contiguous genomic regions that are differentially methylated between type 2 diabetic donors and normoglycaemic controls using WGBS.

Materials and methods

Ethics statement

Written informed consent was obtained from all participants or their relatives. All studies were approved by the local research ethics committees.

Study participants

Study 1

31 sedentary middle-aged men from south of Sweden participated in a 6-months training intervention. Participants were asked to keep their usual lifestyle habits during the study. Overall, adipose tissue biopsies were extracted under the fasting state before and after the intervention from the right thigh and were available from 23 of the participants (**Table 1**).

Table 1. Clinical characteristics of 23 participants with available adipose tissue biopsies in study 1.

Characteristics	Baseline	After exercise	p-value
Age (years)	37.3 ± 4.4		
BMI (kg/m^2)	28.2 ± 2.9	27.9± 3.1	0.18
VO_{2max} (mL/kg/min)	33.1 ±- 4.6	36.2 ± 6.2	0.003
Waist circumference (cm)	97.7 ± 8.6	95.7 ± 8.7	0.02
Diastolic BP (mmHg)	79.3 ±- 9.3	74.8 ± 10.7	0.04
2h OGTT glucose (mmol/L)	6.17 ± 1.02	5.86 ± 1.47	0.32

Data are presented as mean ± standard deviation. Wilcoxon test and two-tailed p-values were used to detect differences between the groups.

Studies 2 and 3

Adipose tissue biopsies for these studies were collected from healthy males, including baseline samples from the cohort recruited for study 1 as well as cohorts from three previously published studies [24,95,96]. The main discovery cohort consisted of 119 samples (**Table 2**), all used in study 3. Out of those, only 96 were available to use for study 2 (**Table 3**). Replication cohort in study 2 consisted of 94 female donors (**Table 3**).

Table 2. Characteristics of 119 Scandinavian men included in studies 2 and 3.

Phenotype	Mean ± SD	Min	1st quartile	Median	3rd quartile	Max
Age (years)	31.03 ± 12.3	22	24	25	35	80
BMI (kg/m^2)	24.91 ± 3.7	16.4	22.2	24.6	27.15	39
HbA1c (%)	4.93 ± 0.48	3.7	4.7	5	5.2	6.4

Data are presented as mean ± standard deviation.

Table 3. Characteristics of 96 men and 94 female donors included in study 2.

Characteristic	Male discovery cohort (n=96)	Female validation cohort (n=94)
Age (years)	32.4 ± 12.8 (23–80)	29.2 ± 4.2 (21–37)
BMI (kg/m^2)	25.6 ± 3.7 (17.5–39.0)	27.2 ± 6.7 (18.2–44.9)
HbA1c (%)	5.3 ± 0.3 (4.7–6.4)	5.0 ± 0.3 (4.4–5.7)

Data are presented as mean ± standard deviation.

Study 4

Human pancreatic islets from 14 cadaveric donors, including 6 donors previously diagnosed with type 2 diabetes, were obtained from the Nordic Network for Islet Transplantation and further processed within the Human Tissue Lab, Lund University Diabetes Centre. The samples were group wise matched for age, sex and islet purity (**Table 4**). These samples are part of a larger cohort of 89 previously described islet donors [71], which was further used in this study for biological and technical replication. The islets were cultured for 4.0±0.2 days prior to DNA and RNA extraction. DNA and RNA from human pancreatic islets were extracted using the AllPrep DNA/RNA Mini Kit (Qiagen GmbH, Hilden, Germany).

Table 4. Characteristics for human donors of pancreatic islets included in the WGBS analysis

	Controls (n=8)	Type 2 diabetic (n=6)	p-value
Sex (m/f)	4/4	3/3	
Age (years)	52.5 ± 3.2 (40-67)	58.2 ± 3.6 (45-66)	0.26
HbA1c (%)	5.47 ± 0.10 (*n*=7; 5.2-6.0)	7.12 ± 0.21 (6.3-7.8)	< 0.0001
BMI (kg/m^2)	24.9 ± 0.3 (23.9-26.6)	28.0 ± 2.0 (22.9-34.6)	0.10

Data are presented as mean ± standard error of the mean. t-tests and two-tailed p-values were used to detect differences between the groups.

Phenotype characterization

Anthropometric traits such as weight, height, waist circumference and waist-hip ratio were measured for all adipose tissue donors. BMI was calculated as a weight in kilograms divided by height in meters squared.

Fasting glucose and fasting insulin were measured before an oral glucose tolerance test (OGTT), which is a two-hour long series of glucose and insulin level measurements performed in individuals after an overnight fast followed by 75g glucose intake.

Average plasma glucose levels over 8 to 12 weeks were characterized as glycated haemoglobin (HbA1c), measured with Mono S method [97].

The homeostatic model assessment of insulin resistance (HOMA-IR) and beta-cell function (HOMA-B) were measured as previously described [98].

For donors included in Study 1, maximal oxygen uptake (VO_{2max}) was assessed on an ergometer bicycle (Ergonomedic 828E, Monark, Sweden), and fat mass was measured with BIA 101 Body Impedance Analyzer (Akern Srl, Pontassieve, Italy).

Genotyping

DNA for genotyping was extracted from whole blood using Gentra Puregene Blood Kit (Qiagen, Hilden, Germany) and analysed using Illumina Human OmniExpress BeadChip and Illumina iScan system. Illumina Human OmniExpress BeadChip is a 24-sample genotyping array that allows for direct genotyping of 731,412 SNPs. Genotype calling was performed using Illumina GenomeStudio® software. Subsequently, we used Plink [99] to identify SNPs with missing values for more than 5% of the interrogated samples, SNPs that deviate from Hardy-Weinberg equilibrium ($p < 0.001$) or low frequency variants

(MAF < 0.05), which were excluded from further analysis. Samples with call rate lower than 98% and samples with mismatched sex label based on X-chromosome heterozygosity were also excluded from further analysis.

Linkage disequilibrium (LD) between SNPs was computed using Trio package from Bioconductor [100].

DNA methylation

Bisulfite conversion

The EZ DNA methylation kit (Zymo Research, Orange, CA, USA) was used for bisulfite conversion prior to Illumina 450k array analyses in studies 1 - 3. For bisulfite conversion of DNA prior to whole-genome sequencing (study 4), the EZ DNA Methylation Gold kit was used.

Illumina 450k BeadChip array

To assess DNA methylation with the Illumina 450k array in studies 1 - 3, the following procedure was used. First, raw methylated and unmethylated intensity values proportional to number of methylated and unmethylated cells in the analysed DNA material were extracted from GenomeStudio®, a proprietary software suite developed by Illumina, and converted to M-values, defined as $log_2\left(\frac{\max(M,0)+1}{\max(U,0)+1}\right)$, where M and U are methylated and unmethylated intensity values of the interrogated CpG site, respectively. M-values were previously suggested as a DNA methylation measure instead of the more commonly used beta-values, defined as a fraction of methylated signal to overall signal: $\frac{\max(M,0)}{\max(U,0)+\max(M,0)+100}$, due to heteroscedasticity in hyper- and hypomethylated CpG sites reported in the latter [85].

Next, for every array the background noise was removed by subtracting the average signal of 614 negative control probes, and quantile normalization [101] was used to correct for intra-array variation.

To correct for probe type bias, we next used the beta-mixture quantile dilation (BMIQ) method, which is a normalization procedure specifically designed to correct for the bias between the two different probe types on the Illumina 450k array [102]. Next, potential sources of batch effects, such as subcohort, sampling

occasion or technical batch were removed using the ComBat normalization procedure [103]. The full analysis pipeline is presented in **Figure 4**.

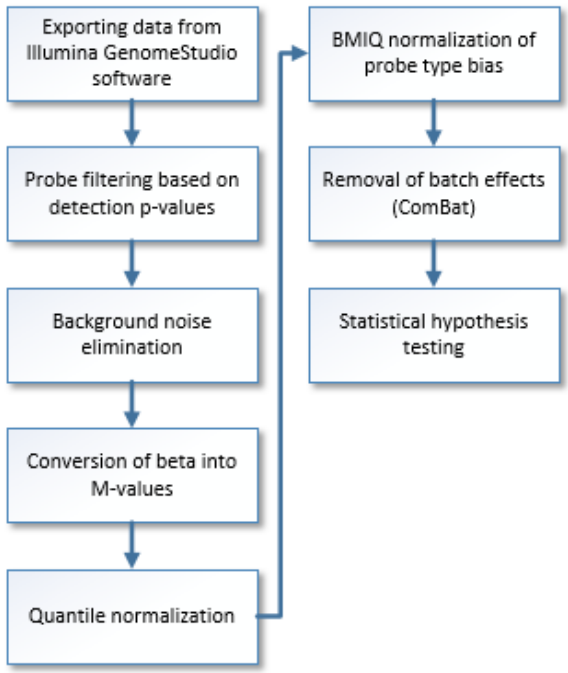


Figure 4. Illumina 450k array analysis pipeline

Pipeline for analysis of Illumina 450k array DNA methylation data, used in studies 1 – 4. First, probes are extracted from the proprietary Illumina software GenomeStudio and low quality probes are filtered based on detection p-values provided by the software. Next, DNA methylation levels are corrected for background noise by subtracting the median levels of negative control probes, which are incorporated into the array. Probes are subsequently converted to M-values and normalized using Quantile normalization. Next, probe type bias is corrected using BMIQ and ComBat is used for batch effect removal.

Whole-genome bisulfite sequencing

Even the most comprehensive microarray technology up to date, the Illumina MethylationEPIC chip, covers only around 3% of the CpG methylome. In order to fully characterize the whole CpG methylome of human pancreatic islets, in study 4, we used next generation whole-genome shotgun sequencing of bisulfite converted DNA.

First, 125-bp long paired end WGBS reads in FASTQ format were generated using Illumina HiSeq2500™. The FASTQ format provides a Phred quality measure, which is defined as $-10\log_{10} P$, where P is a probability of erroneous base call. One of the issues associated with the sequencing by synthesis, utilized by Illumina

HiSeq2500, is that Phred quality drops towards the end of the read [104]. Also, reads can be susceptible to the contamination of sequencing adaptors [105]. In order to tackle both these problems, we used TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore) to trim standard Illumina adaptors as well as low quality bases (Phred < 20) from the 3' end of the reads. Following the trimming procedure, FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to validate read quality.

Next, reads were aligned to the reference genome hg38 using Bismark aligner [106], which internally uses Bowtie 2 [107] for both genome indexing and alignment. DNA methylation was quantified using the Bismark methylation calling algorithm [106]. CpG sites with coverage less than 10 reads were removed from the further analysis. Methylation was calculated as a beta-value, i.e. as a ratio of number of reads aligned to the methylated strand to the total number of reads aligned to the interrogated CpG site.

In order to correct for variability in coverage between consecutive CpG sites, methylation profiles were smoothed using a sliding window approach with the BSmooth package [108].

Next, CpG sites with coverage less than 10x, as well as CpG sites on the negative strand, were filtered from the subsequent analysis, and differentially methylated regions (DMRs) were called using the bsseq package from Bioconductor [108]. A DMR was defined as a set of 3 or more consecutive CpG sites differentially methylated between the groups, with a maximum distance between the CpG sites $\leq$ 300 bp, and with an average absolute difference between groups $\geq$ 5%.

Gene expression

Affymetrix GeneChip® Human Gene 1.0 ST Array chip

Affymetrix GeneChip® Human Gene 1.0 ST Array is a microarray that allows for the robust measurement of mRNA expression profiles of 28,869 coding genes on both exon and whole-transcript level. Transcript level summaries for mRNA expression data were obtained using Robust Multichip Average (RMA) procedure [109] and subsequently batch corrected using ComBat [103].

Whole-genome mRNA sequencing (RNA-seq)

Paired end reads in FASTQ format were generated with the Illumina HiSeq2000 platform. Reads were next aligned to the reference genome hg38 using RSEM software [110], and Gencode v.22 was used for gene annotation. Transcript abundances were quantified as transcript per million (TPM) metric.

Statistical and bioinformatic analysis

Statistical and analysis tools

For the analysis of Illumina 450k array and microarray mRNA expression data from Affymetrix GeneChip® Human Gene 1.0 ST Array, all statistical analyses were performed using R programming language [111]. Differences in mRNA expression or DNA methylation between groups were analysed using Wilcoxon-Mann-Whitney u-test or Student's t-test as implemented in base R functionality. Associations between mRNA expression or DNA methylation and quantitative covariates such as age, BMI or HbA1c were analysed using linear regression models. Parallel calculations were performed on a Condor cluster using R package doMC.

For mQTL analysis, the Matrix EQTL [112] was used, a software that utilizes linear algebra optimisation techniques to perform quick multiple association testing between genotype data and multiple quantitative traits while controlling for covariates.

Multiple hypothesis testing was controlled for using Benjamini-Hochberg False Discovery Rate (FDR) [113], as implemented in standard R function p.adjust.

For mQTL associations, p-values were corrected for multiple hypotheses testing using a modified Bonferroni procedure that takes into account potential linkage between SNPs using LD threshold of $r^2 < 0.9$.

The WGBS analysis pipeline was implemented using Python programming language [114].

Statistical significance of overlap between two different genomic features, i.e. DMRs and functional gene regions such as enhancers, was analysed using a permutation strategy. Regions were shuffled across the genome by randomly and uniformly selecting a new genomic location for each region, while preserving the region length. To build a reliable null distribution, 1,000,000 iterations were performed. The permutation software was implemented using C++.

Results

Study 1:

A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue

In this study we wanted to characterize changes in the DNA methylation pattern in human adipose tissue in response to 6-months exercise exposure. DNA methylation from 31 men, some with a family history of type 2 diabetes, was analysed before and after exercise using the Illumina 450k array. Statistical significance was accessed with t-test and two-tailed p-values.

Overall, we identified 17,975 differentially methylated CpG sites, corresponding to 7,663 RefSeq genes. Out of those, 16,740 sites showed increased DNA methylation in response to exercise, with 911 CpG sites showing an absolute increase of more than 5%. At the same time, 1,505 CpGs demonstrated a decrease in DNA methylation levels, with 98 CpG sites showing more than 5% decrease. Strikingly, differentially methylated CpG sites were overrepresented in intergenic regions and gene bodies while underrepresented in promoter region, defined as a region of 200 bases upstream from transcription start site (TSS200) as well as 1st exon (**Figure 5**). Moreover, differentially methylated CpG sites were underrepresented in CpG islands (**Figure 5**).

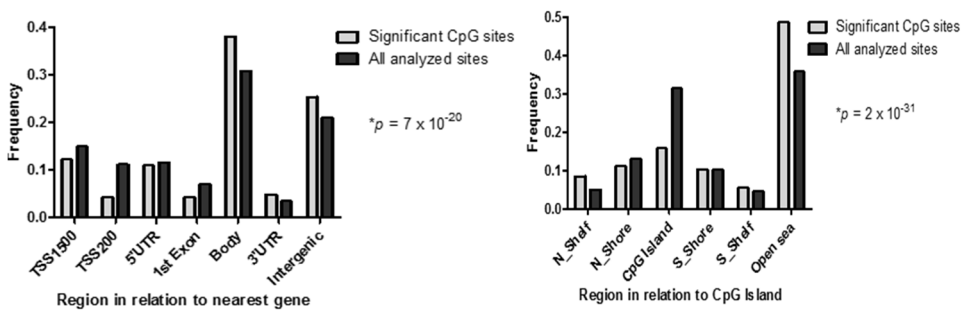


Figure 5. Frequencies of differentially methylated CpG sites in relation to functional gene regions and CpG island sub-regions

* - p-values represent differences between distributions of all analysed 450k array CpG sites comparing to significant sites. Significance was accessed using chi-squared test.

Next, we investigated a potential relationship between DNA methylation and candidate genes for type 2 diabetes and obesity, previously identified in the course of GWASs [115]. Interestingly, 18 out of previously identified 39 type 2 diabetes and 21 out of 53 obesity loci contained differentially (FDR < 0.05) methylated CpG sites in the gene body or in a region 1500kb upstream of TSS.

We further investigated whether the differentially methylated CpG sites were located within genes that also exhibited differential mRNA expression in adipose tissue in response to the 6-months exercise intervention. To analyse gene expression patterns, we used the Affymetrix GeneChip® Human Gene 1.0 ST array. Overall, 197 genes both demonstrated differential mRNA expression and contained at least one differentially methylated CpG site, including *RALBP1*, previously shown to play a role in metabolic syndrome biology [116]. We next used a luciferase expression plasmid to functionally validate that increased DNA methylation of the *RALBP1* promoter suppresses transcriptional activity of the downstream gene ($p < 0.05$).

Study 2:

Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood

For this study we used Illumina Infinium HumanMethylation450 BeadChip and Affymetrix GeneChip® Human Gene 1.0 ST array to quantify DNA methylation and mRNA expression in adipose tissue of 96 male and 94 female donors. Data were modelled as a random effect mixed model using DNA methylation or mRNA expression as a dependent variable, cohort as a random effect variable, and donor characteristics, namely age, BMI and HbA1c, as fixed effect variables. In this study, we used 96 male donors as a discovery cohort, while 94 female donors were used as the replication cohort.

Overall, in the male discovery cohort, we identified 31,567 CpG sites associated with age, 33,058 with BMI and 711 with HbA1c (FDR < 0.05). Some CpG sites were significantly associated (FDR < 0.05) with more than one covariate: 1,334 for both age and BMI, two for age and HbA1c and 12 for BMI and HbA1c. In the female discovery cohort, we found 62 CpG sites associated with age, 39,533 with BMI and seven with HbA1c, yielding overlaps between male discovery and female replication cohorts of 42, 4,979 and 0 CpG sites for age, BMI and HbA1c respectively.

In the mRNA expression analysis, we identified 1,084 transcripts associated with age, 2,936 associated with BMI and 2 with HbA1c.

Interestingly, CpG sites significantly associated with age were overrepresented in CpG islands and underrepresented in the open sea region, while the reverse picture was observed for CpG sites associated with BMI (**Figure 6**).

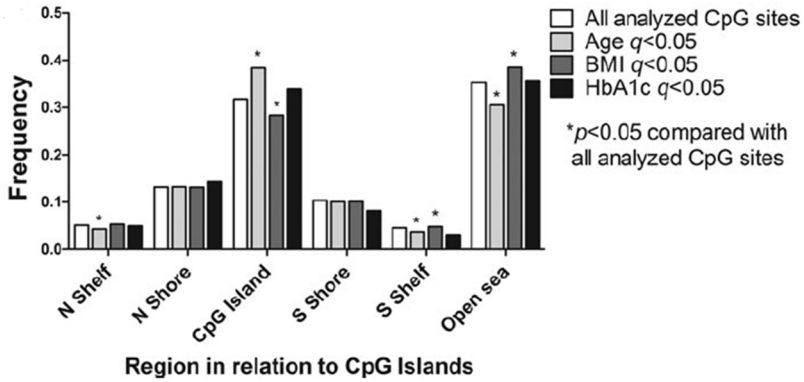


Figure 6. Distribution of CpG sites significantly associated with Age, BMI, HbA1c in adipose tissue by CpG island regions

* - Regions with over- or under-representation of significant CpG sites, accessed using chi-squared test.

Some CpG sites associated with BMI were annotated to previously identified obesity and type 2 diabetes loci, such as *FTO*, *TCF7L2*, *IRS1* and *IRS2*. Moreover, 2,825 out of 2,936 genes which exhibited significant associations between mRNA expression and BMI also contained a BMI-associated CpG site within *cis* distance (500kb upstream from TSS and 100kb downstream from transcription end site (TES)).

We also investigated whether the identified CpG sites are associated with type 2 diabetes by comparing the results in the present study with a previously published case-control cohort of adipose tissue samples from subjects with or without type 2 diabetes [24]. Here, out of 31,567 CpG sites associated with age in the present study, 1,278 sites were also found in the type 2 diabetes case-control cohort. Similarly, 30 out of 711 CpG sites associated with HbA1c as well as 988 out of 33,058 CpG sites associated with BMI were also found to be associated with type 2 diabetes.

Study 3:

A genome-wide mQTL analysis in human adipose tissue identifies genetic variants associated with DNA methylation, gene expression and metabolic traits

The goal of this study was to identify potential interactions between the genome and the epigenome in human adipose tissue by performing a so-called methylation quantitative trait loci (mQTL) analysis, i.e. direct association testing between 592,794 common genetic variants (SNPs) and DNA methylation of 477,891 CpG sites.

Here, we used Illumina Infinium HumanMethylation450 BeadChip to measure DNA methylation and Illumina Human OmniExpress BeadChip for genotyping. Next, the eQTL package for the R programming language was used to perform association testing of SNP-CpG pairs in both *cis* and *trans*, where *cis* distance was defined as 500kb, yielding 112,842,462 *cis* and 283,290,917,454 *trans* SNP-CpG pairs overall. Multiple testing was controlled for using a modified Bonferroni procedure that takes into account the LD structure between genetic variants. Data was modelled as a linear regression with DNA methylation as a dependent variable and SNP, sub-cohort, age and BMI as independent variables.

In *cis,* we identified 101,911 significant SNP-CpG pairs corresponding to 51,143 unique SNPs and 15,208 unique CpG sites, annotated to 5,589 genes (**Table 5**). These include previously reported GWAS loci associated with obesity, such as *POMC/ADCY3* [117], lipid profiles, such as *CETP* [118], and fasting glucose, such as *ACADS* [119].

In trans, we identified 5,342 significant SNP-CpG pairs corresponding to 2,735 unique SNPs and 596 unique CpG sites, annotated to 375 genes (**Table 5**), including PTBP2, which was previously reported to be associated with BMI [117].

Table 5. Number of significant mQTLs identified in human adipose tissue.

	cis-mQTL	trans-mQTL
SNP-CpG pairs	101,911	5,342
SNPs	51,143	2,735
CpG sites	15,208	596
Unique genes	5,589	375

Interestingly, significant *cis*-SNP-CpG pairs were overrepresented on chromosome 6, 7, 8, 13 and 21, with the highest deviation from expected frequency on chromosome 6 (p-value=3.4×10^{-89}) (**Figure 7**).

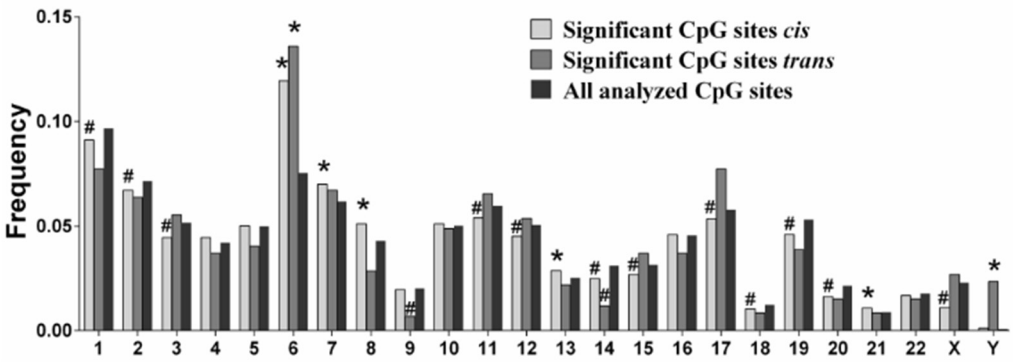


Figure 7. Chromosomal distribution of CpG sites of significant mQTLs

* - Chromosomes with over-representation of significant CpG sites
# - Chromosomes with under-representation of significant CpG sites
Significance of over- or under-representation was assessed using chi-squared test.

The mQTL analysis was followed up by performing expression quantitative trait loci (eQTL) analysis for SNPs of significant mQTLs in both *cis* and *trans*. For the 51,143 significant *cis*-mQTL SNPs, we identified 926 SNP-mRNA transcript pairs, corresponding to 635 unique SNPs and 101 unique mRNA transcripts (**Table 6**). For the 2,735 significant *trans*-mQTL SNPs, we identified 89 SNP-mRNA transcript pairs corresponding to 89 unique SNPs and 14 mRNA transcripts (**Table 6**).

Table 6. Number of significant eQTLs identified in human adipose tissue.

	eQTLs of *cis*-mQTL-SNPs	eQTLs of *trans*-mQTL-SNPs
SNP-mRNA transcript pairs	926	270
Unique SNPs	635	89
Unique mRNA transcripts	101	14
Unique genes	86	10

Amongst the mRNA transcripts uncovered in the *cis* m/eQTL analysis were *L27A* and *THNSL2*, previously linked to obesity [120], and *G6PC2*, associated with glycaemic traits [121].

Next, we used the causal inference test (CIT) [122] to investigate the causal relationship between DNA methylation, genotype and metabolic phenotypes such as BMI, fasting glucose, fasting insulin, HOMA-B, HOMA-IR, HbA1c, cholesterol, triglycerides and high (HDL) and low (LDL) density lipoproteins. Despite having relatively low power for this type of analysis, we identified 35 SNPs that were associated with a metabolic phenotype in such a way that DNA methylation plays a direct mediating role between this SNP and the phenotype (**Table 7**).

Table 7. Causal Inference Test results.

Chr	CpG ID	CpG Gene	SNP ID	Phenotype	CIT causal p-value
6	cg12929486	*SLC22A16*	*rs2428190*	BMI	0.02
5	cg14825688	*LEAP2*	*rs39830*	Fasting glucose	0.03
5	cg14825688	*LEAP2*	*rs803217*	Fasting glucose	0.03
2	cg01726273		*rs4853438*	Fasting insulin	0.03
8	cg11123440		*rs12458*	HOMA-B	0.03
12	cg10240950	*C12orf76*	*rs1027949*	HOMA-IR	0.05
12	cg10240950	*C12orf76*	*rs10774978*	HOMA-IR	0.05
12	cg10240950	*C12orf76*	*rs11068984*	HOMA-IR	0.05
10	cg26169081	*CAMK1D*	*rs11257926*	HOMA-IR	0.04
10	cg26169081	*CAMK1D*	*rs17152029*	HOMA-IR	0.01
10	cg26169081	*CAMK1D*	*rs17152037*	HOMA-IR	0.04
12	cg10240950	*C12orf76*	*rs2302689*	HOMA-IR	0.05
7	cg17372657		*rs1880296*	HbA1c	0.03
7	cg17372657		*rs2949170*	HbA1c	0.03
7	cg17372657		*rs2949192*	HbA1c	0.03
6	cg13561028	*SFTA2*	*rs3130782*	HbA1c	0.01
6	cg13561028	*SFTA2*	*rs3131934*	HbA1c	0.04

Study 4:

Whole-genome bisulfite sequencing of human pancreatic islets reveals novel differentially methylated regions in type 2 diabetes

In this study, the DNA methylation landscape of human pancreatic islets from 6 diabetic donors and 8 normoglycaemic controls was investigated using WGBS. We identified 25,820 DMRs, out of which 13,696 demonstrated higher and 12,124 lower DNA methylation in islets from type 2 diabetic compared to control donors. Strikingly, 55% of the identified DMRs were located within a region spanning 50kb upstream of TSS. Several DMRs were located in a close proximity to type 2 diabetes candidate genes identified by the GWAS studies, such as *PDX1*, *GLIS3*, *THADA*, *KCNQ1* and *TCF7L2* [46]. The DNA methylation profile of a DMR in the promoter of *PDX1,* which encodes for a transcription factor previously implicated in islet function [123], is depicted in **Figure 8**.

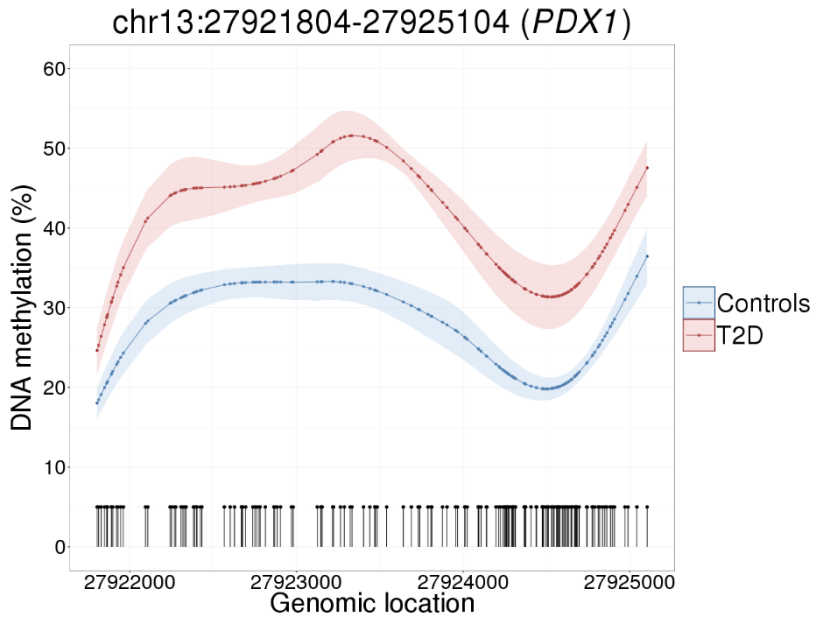


Figure 8. DMR located in the promoter region of *PDX1*

This figure depicts a ~3000 bp long DMR located on chromosome 13 and covering a promoter of *PDX1*. Lollypops represent each CpG site in the DMR. Areas around the line connecting DNA methylation values depict intervals of ± 1 standard error of the mean.

We further looked into a potential involvement of identified DMRs into the mechanism of transcriptional regulation in human pancreatic islet cells. Interestingly, binding sites of transcription factors previously implicated in the islet function, such as FOXA2, MAFB, NKX2.2, NKX6.1 and PDX1, were statistically overrepresented in the DMRs (all $p < 1*10^{-6}$). Interestingly, binding sites of transcription factor NKX2.2 overlapped with more than 7.5% of the DMRs. Moreover, 12,911 (49.8%) of the identified DMRs overlapped with at least one of the ~700.000 cross-tissue dynamic DMRs previously identified by Ziller et al. [59].

Next, we analysed mRNA expression data for the same 14 individuals using RNA-seq and separated 60,483 Gencode transcripts [124] into 4 groups according to their TPM values: not-expressed transcripts (n = 38,261), transcripts with low TPM value (n = 7,407), medium TPM value (n = 7,407) and high TPM value (n = 7,408). Non-expressed genes showed significantly higher DNA methylation in and around the promoter region and the first exon, but lower methylation levels in all exons except first, intergenic regions and the region of 10kb downstream of TES (**Figure 9**).

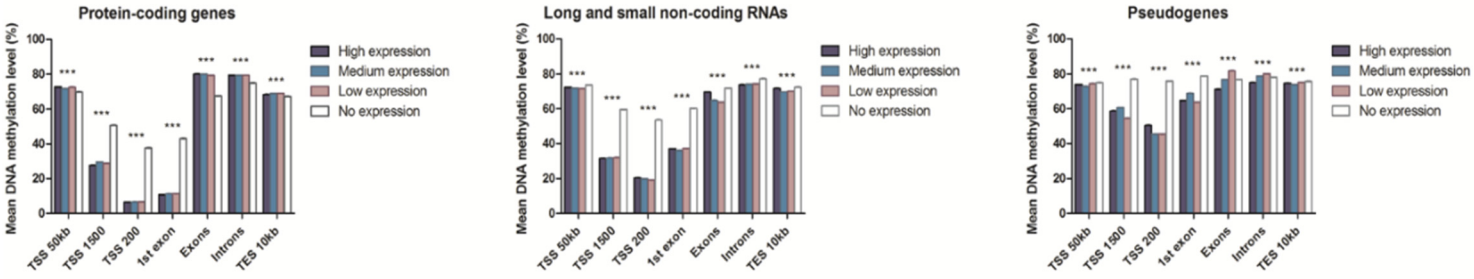


Figure 9. DNA methylation levels of different gene regions

Average DNA methylation levels for different transcript types, separated by genomic regions. TSS 50kb represents 1,501-50,000bp upstream from the transcription start site, TSS 1500 represents 201-1,500bp upstream from the transcription start site, TSS 200 represents 1-200bp upstream from the transcription start site and TES 10kb represents 1-10,000bp downstream of the transcription end site.

We further looked into the genomic distribution of the DNA methylation in human pancreatic islets on a single CpG site resolution. Using the Gencode annotation of the human transcriptome, version 22 [124], we extracted genomic locations for the following regions of every gene: 50kb upstream from TSS, 1500bp upstream of TSS, 200bp upstream of TSS, 1st exon, all exons except 1st, introns, 10kb downstream from TES and looked into average DNA methylation levels for these regions. Interestingly, we noticed a variability in DNA methylation patterns between three different Gencode transcript types, namely protein-coding transcripts, long and small non-coding RNA transcripts and pseudogenes. While regions in close proximity to TSS in protein-coding transcripts were hypomethylated, they exhibited medium methylation levels for non-coding RNAs and hypermethylation for pseudogenes (**Figure 10**).

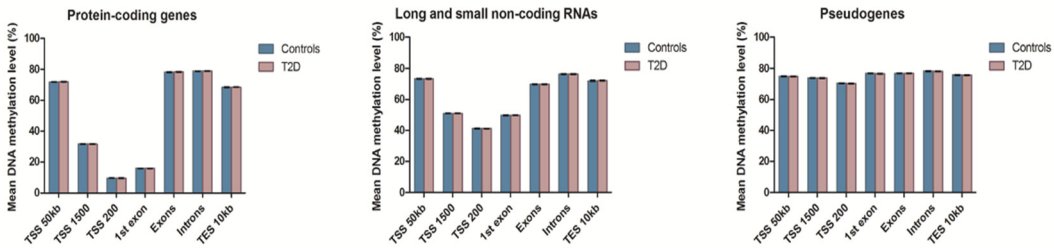


Figure 10.

DNA methylation levels for different Gencode transcript types: protein-coding genes, long and small non-coding RNAs, and pseudogenes.

Discussion

DNA methylation is thought to be one of the mechanisms by which the environment can affect cellular function. Here, we presented 4 studies where we performed DNA methylation profiling of two tissues known to play an important role in the pathogenesis of type 2 diabetes, namely pancreatic islets and adipose tissue.

In study 1, we have shown that DNA methylation in human adipose tissue can exhibit dynamic properties in response to a 6-months exercise intervention. The effect of exercise on the human methylome was previously shown in skeletal muscle [74], but whether it affects DNA methylation in adipose tissue remained unknown. Overall, we discovered 17,975 CpG sites with statistically different methylation levels between individuals before and after exercise. Interestingly, the majority (92%) of the significant CpG sites increased DNA methylation levels, and most of the identified CpG sites were located in gene bodies or intergenic regions. Increased DNA methylation in gene bodies was previously found to be positively correlated with gene expression, and it was hypothesized that DNA methylation is involved in transcriptional elongation and alternative splicing events [58].

Next, we found that CpG sites significantly different before compared with after the intervention were statistically underrepresented in CpG islands, which is in agreement with other studies and can be explained by the protection of CpG islands from DNA methylation, which was later supported by an experimental study [125]. Interestingly, we observed alterations in DNA methylation in CpG sites located in or near genes previously associated with diabetes or obesity-related traits. For example, differentially methylated CpG sites were located in the gene body as well as promoter region of *ITPR2* gene, previously associated with waist-hip ratio [126], and in the gene body of the type 2 diabetes candidate gene *TCF7L2*. This study potentially suggests that exercise affects metabolism and cellular function through altered gene activity or genome accessibility.

In study 2 we showed that intra-individual differences in age, BMI and HbA1c, which are all known risk factors of type 2 diabetes and obesity, can be reflected in the human adipose tissue methylome. The study was performed in a discovery cohort of 96 human male and replicated in a cohort of 94 female adipose tissue donors.

Overall, ~ 5000 CpG sites were associated with BMI in both discovery and replication cohorts. Interestingly, a large fraction of those was also found in the case-control cohort of diabetic and non-diabetic adipose tissue donors, pointing not only to the fact that differences in BMI can largely affect the adipose tissue methylome, but also that these differences can point to type 2 diabetes predisposition, linking obesity and type 2 diabetes through epigenetic changes. We have also tested whether DNA methylation of identified sites was associated with mRNA expression in human adipose tissue, identifying ~3000 genes, including many which were previously linked to obesity in GWAS, such as *FTO*, *TCF7L2*, *IRS1* and *IRS2*.

Despite the fact that almost ~30,000 CpG sites were significantly associated with age in the discovery cohort, only 42 of those were also identified in the replication cohort. This is most likely explained by the modest age span in the replication cohort of 16 years, compared to the almost 60-year age span in the male discovery cohort. Anyhow, many of the genes covered by the 30,000 CpG sites identified were previously both linked to the aging processes and shown to exhibit age-related epigenetic changes in DNA methylation in blood and pancreatic islets, including genes such as *ELOVL2*, *FHL2*, *KLF14* and *GLRA1* [127,128].

No CpG sites were associated with HbA1c in both the discovery and replication cohorts. This can potentially be due to a relatively small span in HbA1c in both male and female cohorts, sex differences or to the fact that glucose levels are not really reflected in the adipose tissue epigenome at all. However, larger validation cohorts specifically targeting glycaemic traits are required to prove or disprove either hypothesis.

The goal of study 3 was to unravel potential relationships between the genome and methylome in human adipose tissue. The fact that DNA methylation is affected by genetic variation was supported by a recent experimental study [125]. Here, to characterize the properties of the adipose tissue methylome that can potentially be under genetic control, we constructed a map of interactions between SNPs and CpG methylation in both *cis* and *trans*. In previous studies, majority of the identified SNP-CpG interactions occur in *cis* [80], and in line with this results we discovered 101,911 statistically significant SNP-CpG pairs in *cis* and only 5,342 SNP-CpG pairs in *trans*. Moreover, most of the SNP-CpG pairs identified in *cis* were located close to each other. The number of significant mQTLs in both *cis* and *trans* characterized in this study is in line with previous mQTL study in pancreatic islets, and the same can be said about a distribution of distances between identified SNPs and CpG sites [80].

We found an underrepresentation of *cis* mQTLs in CpG islands and gene promoters, while at the same time, *cis* mQTLs were overrepresented in gene bodies and intergenic regions. These findings not only point to a conserved

function of CpG islands, but also imply that CpG sites affected by genetic factors tend to be located outside of the promoters, meaning that identified SNPs may act on transcriptional elongation or alternative splicing rather than transcriptional initiation. Significant mQTLs were also overrepresented on chromosome 6, probably due to the HLA region, which is prone to genetic variation often associated with immunological and metabolic disorders [129,130].

Interestingly, many of the genes previously identified in metabolic and lipid traits GWAS, such as *POMC, GIPR, GRB10, FADS2, SORT1* and *APOA5*, were located in *cis* distance to identified adipose tissue mQTLs. Moreover, many of the identified mQTL SNPs were also directly found in GWAS consortia studies [117–119,126,131,132], while some of the identified SNPs correlated with metabolic phenotypes in the study cohort. Taken together, this point to the interplay between DNA methylation and genetic variation in adipose tissue to be involved in the regulation of metabolic traits.

Pancreatic islets have a key role in the pathogenesis of type 2 diabetes. Previous studies have identified a set of CpG sites associated with type 2 diabetes in pancreatic islets using the Illumina 450k or 27k BeadChip arrays [71,133]. However, these arrays cover at most ~1.5% of the human DNA methylome. Due to the higher coverage in comparison to the Illumina 450k BeadChip array, WGBS allows for a more detailed and precise interrogation of the dynamic properties of the methylome, especially in the genomic regions poorly covered by the array, such as the enhancers. Therefore, in study 4 we used WGBS technology to profile ~24 million CpG sites in human pancreatic islets from 14 donors with or without type 2 diabetes.

Previously, we have shown that absolute differences in DNA methylation between islet donors with and without type 2 diabetes are relatively small (< 10%), which would require a coverage of at least 10 reads per CpG site when using WGBS [71]. Moreover, a recent study by Ziller et al. investigated coverage requirements for WGBS and suggested that higher sequencing depth facilitates the detection of shorter DMRs with smaller methylation differences [134]. Thus, we designed the study to have an average coverage of approximately 20X for each CpG sites, in the same time filtering away CpG sites with less than 10X coverage.

Here, we succeeded in identifying 25,820 contiguous genomic regions differentially methylated between normoglycaemic and diabetic donors. Interestingly, ~25% of the DMRs associated with type 2 diabetes were located in the gene bodies, and ~55% within a region upstream of proximal promoters, defined as 1,501 – 50,000 bp from TSS. Moreover, 80% of the DMRs located in the gene bodies were found in intronic regions, which is in line with previous study in brain [135]. Overall, this result, similar to other studies in this thesis, suggests that epigenetic modifications may interact with the cellular function

through transcriptional elongation and differential splicing events rather than by methylation of promoter regions, although future studies are needed to investigate whether intronic DMRs have a special role in alternative splicing events or this result is consequential to the fact that introns are much larger in size comparing to exons.

DNA methylation in enhancer regions were previously shown to have dynamic properties between various cell types [59]. It was also recently shown that dynamic methylation of enhancer regions is a risk factor for mortality in cancer patients [136]. In support of the hypothesis that type 2 diabetes can be reflected in cellular function through enhancer methylation, we found an overrepresentation of identified DMRs in the enhancer regions previously published by Pasquali et al. [137].

PDX1 is expressed in pancreatic islets, encoding a transcription factor known to play a role in islet development as well as in regulation of insulin expression in mature beta-cells [138]. It was shown that in mice *PDX1* silencing can cause type 2 diabetes [139]. Additionally, increased DNA methylation levels in distal promoter and enhancer regions of *PDX1* was previously linked to decreased *PDX1* expression [140]. In line with this study, two DMRs with the highest overall differences in DNA methylation among all 25,820 identified DMRs, and 7 DMRs overall, were located in or close to the *PDX1* gene, covering both enhancer and promoter regions. This result implies that epigenetic regulation of *PDX1* expression plays a major role in regulating cellular function of human pancreatic islets.

In study 3, we displayed that interactions between SNPs and CpG methylation can play an important role in adipose tissue biology. The importance of analysing genetic variation in its surrounding epigenetic context was recently shown in the study investigating variants in the *FTO* gene [49]. These observations were further explored in study 4, where we found rs163184, a GWAS-identified variant in the *KCNQ1* gene, to be located in one of the identified DMRs, while several other GWAS-identified loci, such as *GLIS3*, *THADA*, *KCNQ1, TCF7L2* and *ADCY5*, were among the genes with the highest number of annotated DMRs. This supports a potential role of DNA methylation in the interplay between genetic variation in named loci and cellular and metabolic function.

Cell-type heterogeneity is an important issue in the analysis of DNA methylation in composite tissues such as pancreatic islets or adipose tissue [141]. Methods to adjust DNA methylation values for cell-type composition utilize either a prior knowledge about DNA methylation levels in individual cell-types [142], or are reference free, such as methods proposed by Houseman et al. or Zou et al. [72,142]. However, the mathematical assumptions underlying the reference free methods were not thoroughly evaluated until recently [141].

Thus, in the study 1, we evaluated whether cell-type heterogeneity may be affected by exercise by analysing mRNA expression of genes known to be specific to adipocytes, preadipocytes, macrophages and cytokines. Our results support that no significant differences in cell-type composition occur before versus after exercise intervention. We have further shown that there are no significant differences in beta- and alpha-cell composition in pancreatic islets from donors with type 2 diabetes compared with non-diabetic controls [71], supporting that cell composition does not affect our results in study 4. Moreover, we overlapped the identified DMR set in study 4 with cell-type specific DMRs previously identified by Ziller et al. [59]. Here, almost 50% of T2D-related DMRs overlap with at least one genomic region previously shown to exhibit differences in DNA methylation between multiple cell-types and it is possible that these genomic regions are particularly dynamic to epigenetic changes. However, further studies are needed to fully characterize the epigenome in individual cell types from human tissues.

Summary and general conclusion

Both pancreatic islets and adipose tissue play an important role in metabolic processes in humans. In this thesis we have shown that the DNA methylome in these tissues reflects environmental factors, metabolic phenotypes and type 2 diabetes status. Next, we have shown that regions of the methylome associated with these factors tend to occur more frequently in some genomic features, such as gene bodies and intergenic regions, than others, and often tend to be under genetic control. Overall, during the work presented in this thesis, we performed 4 studies.

In study 1, we characterized changes that the DNA methylome in adipose tissue exhibit in response to 6-months exercise intervention and show that these changes occur in or close to several genes previously implicated in type 2 diabetes pathogenesis.

In study 2, we show that DNA methylation in human adipose tissue is associated with age, BMI and HbA1c.

In study 3, we characterize the relationship between epigenetic and genetic variation in human adipose tissue. Here, we show that many of the SNPs previously identified to be involved in lipid biology or glucose metabolism tend to be interacting with DNA methylation in either *cis* or *trans*.

In study 4, we provide a map of the DNA methylation landscape in human pancreatic islets and identify genomic regions that exhibit dynamic properties in relation to type 2 diabetes. We find a non-random occurrence of such regions in genes related to type 2 diabetes, as well as enhancer regions and gene bodies. This map can be an important reference for future studies of genetics or epigenetics of type 2 diabetes in order to localize causal variation.

Despite the fact that these studies characterize the DNA methylome in adipose tissue and pancreatic islets in relation to type 2 diabetes, age, environmental exposure and metabolic phenotypes, future studies are required to experimentally investigate mechanisms of interaction and cause-effect relationship between these factors, epigenetics and genetic variation.

Popular summary

Every single living cell keeps all the information about how it should function in a very long DNA molecule, tightly packed inside the cell nucleus. DNA is shaped as a chain of small molecules, each belonging to one of 4 possible types, adenine, thymine, cytosine, and guanine, often referred by their first letters - A, T, C, and G. Combination of this 4 ‘letters’ define genes – direct instructions for the cell how to function. Every cell in every organism contains exactly the same copy of the DNA molecule. It is in the form of DNA hereditary information is passed from generation to generation.

However, this raises several questions. If every cell has exactly the same copy of the DNA, why are there different cells in human body? Next, if the cell function is hardcoded into the DNA, how do human cells respond to environmental stimuli? Why do cells in patients with a disease act differently than in healthy people? One answer to this question is that genes can be turned on and off by molecules called transcription factors. However, transcription factors alone do not explain the whole picture of gene regulation. Another explanation to the gene regulation problem is called epigenetics. Epigenetics is defined as a chemical modification of a DNA molecule ‘above’ the actual DNA code - for example, an attachment of some small other molecules or proteins to it, and studies have discovered that in humans these epigenetic modifications occur everywhere throughout the genome. This epigenetic code, called epigenome (or ‘above’ genome) represents an additional layer of the information which is encoded in the DNA.

Type 2 diabetes is rapidly increasing its prevalence. It was estimated that in year 2040, 624 million people worldwide will carry this disease. Type 2 diabetes is defined as an inability to produce enough insulin, a hormone responsible for reducing the amount of sugar in the blood and for target tissues such as the liver and muscle to respond to insulin and take up glucose. If not absorbed by the tissues, blood sugar is toxic and can damage multiple organs in the body. Currently, what exactly causes type 2 diabetes is not very well understood, but involves heritable as well as life-style related factors.

We know that factors such as age, physical inactivity and a high-calorie diets greatly increase the risk of type 2 diabetes. It is also known that people with a family history of the disease are more likely to get in themselves. Unfortunately,

mechanisms behind this phenomenon are unknown, and up to now scientists have not managed to find all the genes that can explain all the risk for type 2 diabetes.

Epigenetics is proposed to be one of the mechanisms of how cells react to the environment in regard to type 2 diabetes, and the goal of this thesis was to study whether there is a relationship between type 2 diabetes (or underlying factors of type 2 diabetes, such as lifestyle, age, predisposition to obesity etc.) and the epigenome of human tissues. We studied pancreatic islets and fat, which are both known to be involved in the pathogenesis of type 2 diabetes.

Here, we first studied a group of 23 men who had to change their lifestyle for 6 months by exercising 3 times a week. Next, we extracted samples of fat tissues from these men and studied the epigenome in this tissue before and after the exercise intervention. Our results show that the epigenome in fat from these men before differs from the epigenome after the exercise intervention. Moreover, we found that these changes are affecting the function of the genes attributed to the biology of type 2 diabetes and obesity. Next, in study 2, we also show that the epigenome in fat tissue strongly reflects obesity status, blood sugar and age, regulating the functional ability of some very important genes for human metabolism. In study 3, we prove that the epigenome in fat tissue is tightly inter-linked with the DNA sequence and that the epigenome can interact with the genetic code and together affect cell biology. Finally, in study 4 we showed large differences in the epigenome of human pancreatic islets, a tissue responsible for secreting insulin, between donors with and without type 2 diabetes. This is the first time anyone could manage to study almost 90% of the epigenome in this human tissue. We also performed experiments in laboratory cell lines to show how these differences of the epigenome may affect insulin secretion.

In conclusion, in this thesis we show the importance of studying epigenetics in relation to type 2 diabetes, lifestyle and obesity, and that these factors can affect cell function through epigenetic mechanisms in humans.

Acknowledgements

I would like to thank all amazing people in Lund, Copenhagen, Boston and other parts of the globe whom I met during my life as a student of human biology. It was a great honor to work with all of you.

First of all, I would like to thank my supervisor Charlotte Ling. Charlotte, you are the best supervisor one could possibly wish for – smart, hardworking, passionate about your job, and, most importantly, able to transfer your enthusiasm to those around you when it is needed the most. You have taught me to never give up.

Many thanks to my co-supervisor Tina Rönn for being so supportive, smart, attentive to detail, and for all the encouragement and great scientific discussions we had during the time.

Thanks to all my wonderful colleagues in Epigenetics and Diabetes group: Anders for all the football as well as non-football talks, Elin for jokes and great atmosphere in our room, Emma for always being helpful and friendly, Cajsa for being such a considerate and clever person, Kalle for craft beers and football games, Mahboubeh for being so good-natured, Tasnim for your great knowledge, Sonya for your endless energy and for teaching me basic Spanish, Kerstin for making sure we get up from our tables for some beach volleyball and to my best mate Alexander for all the fun we had together during these years.

Many thanks to people in Copenhagen: Allan Vaag and others for creating the great research group which we had so many collaborations with. Special thanks to my friends Line Hjort and Linn Gillberg for not telling anyone about what happened on the whale watching tour in Iceland.

Many thanks to Leif Groop, for it is largely because of him our center exists in it's current state, to Peter Almgren and Claes Landevall with help and advice on statistics and genetics, to Johan Hultman and Mattias Borell for their great help and computational support. I would also like to thank all those people with whom I had a pleasure to interact with during my time in Malmö: Paul Franks, Nikolay Oskolkov, Dmitry Shungin, Lena Eliasson, Petter Storm, Petter Vikman, Valeria Lyssenko, Ulrika Blom-Nilsson, Claes Wollheim, Ulrika Krus, Ola Hansson, and many others.

Thanks to all the great people I met in MIT, and especially to Manolis Kellis for the opportunity to experience this great environment in the frontlines of modern science.

I would like to specially thank Linn Gillberg for the design of the book cover.

Thanks to all the friends in Moscow: Kolyan, Vovan, Pit and many others, those who visited me during the time in Sweden and those who didn't.

Thanks to my mother, father and little sisters back in Russia, for the excessive amount of kicks, love and support throughout my life, nothing would have ever happened without you. To my grandparents for not only teaching me about the most important things in life, but also for making me passionate about science.

And finally, to the most important person in my life, Lena, thank you for all the laughs, coffee, and choosing to go on this adventure together. Thank you for being there for me even at the turbulent times of searching for myself and for always pushing me further with your wisdom and ridiculous optimism. You are my inspiration. Hold on, we've got even more fun ahead!

References

1. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era — concepts and misconceptions. Nat Rev Genet. 2008;9: 255–266. doi:10.1038/nrg2322

2. Kaprio J, Tuomilehto J, Koskenvuo M, Romanov K, Reunanen A, Eriksson J, et al. Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. Diabetologia. 1992;35: 1060–1067.

3. Walley AJ, Blakemore AIF, Froguel P. Genetics of obesity and the prediction of risk for health. Hum Mol Genet. 2006;15: R124–R130. doi:10.1093/hmg/ddl215

4. Barrett JH. Measuring the effects of genes and environment on complex traits. Methods Mol Med. 2008;141: 55–69.

5. Ali O. Genetics of type 2 diabetes. World J Diabetes. 2013;4: 114–123. doi:10.4239/wjd.v4.i4.114

6. Zimmet PZ, Alberti KGMM. Epidemiology of Diabetes—Status of a Pandemic and Issues Around Metabolic Surgery. Diabetes Care. 2016;39: 878–883. doi:10.2337/dc16-0273

7. Uusitupa M. Lifestyles Matter in the Prevention of Type 2 Diabetes. Diabetes Care. 2002;25: 1650–1651. doi:10.2337/diacare.25.9.1650

8. World Obesity Federation. Global prevalence of Adult Obesity, November 2015 [Internet]. [cited 9 Jul 2016]. Available: http://www.worldobesity.org/site_media/library/resource_images/Global_prevalence_of_Adult_Obesity_November_2015_WO_25.11.15.pdf

9. Association AD. Economic costs of diabetes in the U.S. in 2012. Diabetes Care 2013;36:1033–1046. Diabetes Care. 2013;36: 1797–1797. doi:10.2337/dc13-er06

10. American Diabetes Association. 2. Classification and Diagnosis of Diabetes. Diabetes Care. 2015;38: S8–S16. doi:10.2337/dc15-S005

11. Wilcox G. Insulin and Insulin Resistance. Clin Biochem Rev. 2005;26: 19–39.

12. Steiner DJ, Kim A, Miller K, Hara M. Pancreatic islet plasticity: Interspecies comparison of islet architecture and composition. Islets. 2010;2: 135–145.

13. Hauge-Evans AC, King AJ, Carmignac D, Richardson CC, Robinson ICAF, Low MJ, et al. Somatostatin Secreted by Islet δ-Cells Fulfills Multiple Roles as a Paracrine Regulator of Islet Function. Diabetes. 2009;58: 403–411. doi:10.2337/db08-0792

14. Asakawa A, Inui A, Yuzuriha H, Ueno N, Katsuura G, Fujimiya M, et al. Characterization of the effects of pancreatic polypeptide in the regulation of energy balance. Gastroenterology. 2003;124: 1325–1336.

15. Verhulst P-J, Depoortere I. Ghrelin's second life: From appetite stimulator to glucose regulator. World J Gastroenterol. 2012;18: 3183–3195. doi:10.3748/wjg.v18.i25.3183

16. Cerf ME. Beta Cell Dysfunction and Insulin Resistance. Front Endocrinol (Lausanne). 2013;4. doi:10.3389/fendo.2013.00037

17. Donath MY, Halban PA. Decreased beta-cell mass in diabetes: significance, mechanisms and therapeutic implications. Diabetologia. 2004;47: 581–589. doi:10.1007/s00125-004-1336-4

18. Kaiser N, Leibowitz G, Nesher R. Glucotoxicity and beta-cell failure in type 2 diabetes mellitus. J Pediatr Endocrinol Metab. 2003;16: 5–22.

19. Forbes JM, Cooper ME. Mechanisms of Diabetic Complications. Physiological Reviews. 2013;93: 137–188. doi:10.1152/physrev.00045.2011

20. Meneghini LF. Early Insulin Treatment in Type 2 Diabetes. Diabetes Care. 2009;32: S266–S269. doi:10.2337/dc09-S320

21. Bell CG, Walley AJ, Froguel P. The genetics of human obesity. Nat Rev Genet. 2005;6: 221–234. doi:10.1038/nrg1556

22. Guilherme A, Virbasius JV, Puri V, Czech MP. Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. Nat Rev Mol Cell Biol. 2008;9: 367–377. doi:10.1038/nrm2391

23. Scherer PE. Adipose Tissue. Diabetes. 2006;55: 1537–1545. doi:10.2337/db06-0263

24. Nilsson E, Jansson PA, Perfilyev A, Volkov P, Pedersen M, Svensson MK, et al. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. Diabetes. 2014; doi:10.2337/db13-1459

25. Kershaw EE, Flier JS. Adipose tissue as an endocrine organ. J Clin Endocrinol Metab. 2004;89: 2548–2556. doi:10.1210/jc.2004-0395

26. Rutkowski JM, Stern JH, Scherer PE. The cell biology of fat expansion. J Cell Biol. 2015;208: 501–512. doi:10.1083/jcb.201409063

27. Coelho M, Oliveira T, Fernandes R. Biochemistry of adipose tissue: an endocrine organ. Arch Med Sci. 2013;9: 191–200. doi:10.5114/aoms.2013.33181

28. Coll AP, Farooqi IS, O'Rahilly S. The Hormonal Control of Food Intake. Cell. 2007;129: 251–262. doi:10.1016/j.cell.2007.04.001

29. Yadav A, Kataria MA, Saini V, Yadav A. Role of leptin and adiponectin in insulin resistance. Clin Chim Acta. 2013;417: 80–84. doi:10.1016/j.cca.2012.12.007

30. Margetic S, Gazzola C, Pegg GG, Hill RA. Leptin: a review of its peripheral actions and interactions. Int J Obes Relat Metab Disord. 2002;26: 1407–1433. doi:10.1038/sj.ijo.0802142

31. Minokoshi Y, Toda C, Okamoto S. Regulatory role of leptin in glucose and lipid metabolism in skeletal muscle. Indian J Endocrinol Metab. 2012;16: S562–S568. doi:10.4103/2230-8210.105573

32. Cusi K. The role of adipose tissue and lipotoxicity in the pathogenesis of type 2 diabetes. Curr Diab Rep. 2010;10: 306–315. doi:10.1007/s11892-010-0122-6

33. Hajer GR, van Haeften TW, Visseren FLJ. Adipose tissue dysfunction in obesity, diabetes, and vascular diseases. Eur Heart J. 2008;29: 2959–2971. doi:10.1093/eurheartj/ehn387

34. Zhou Y, Rui L. Leptin signaling and leptin resistance. Front Med. 2013;7: 207–222. doi:10.1007/s11684-013-0263-5

35. International HapMap Consortium. The International HapMap Project. Nature. 2003;426: 789–796. doi:10.1038/nature02168

36. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526: 68–74. doi:10.1038/nature15393

37. Hayes B. Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). Methods Mol Biol. 2013;1019: 149–169. doi:10.1007/978-1-62703-447-0_6

38. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. Eur J Hum Genet. 2016; doi:10.1038/ejhg.2015.269

39. Tillil H, Köbberling J. Age-corrected empirical genetic risk estimates for first-degree relatives of IDDM patients. Diabetes. 1987;36: 93–99.

40. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet. 2000;26: 163–175. doi:10.1038/79876

41. Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006;38: 320–323. doi:10.1038/ng1732

42. Prasad RB, Groop L. Genetics of Type 2 Diabetes—Pitfalls and Possibilities. Genes (Basel). 2015;6: 87–123. doi:10.3390/genes6010087

43. Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? Annals of the New York Academy of Sciences. 2010;1212: 59–77. doi:10.1111/j.1749-6632.2010.05838.x

44. Walley AJ, Asher JE, Froguel P. The genetic contribution to non-syndromic human obesity. Nat Rev Genet. 2009;10: 431–442. doi:10.1038/nrg2594

45. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518: 187–196. doi:10.1038/nature14132

46. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44: 981–990. doi:10.1038/ng.2383

47. Taşan M, Musso G, Hao T, Vidal M, MacRae CA, Roth FP. Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. Nat Meth. 2015;12: 154–159. doi:10.1038/nmeth.3215

48. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: Illuminating the Dark Road from Association to Function. Am J Hum Genet. 2013;93: 779–797. doi:10.1016/j.ajhg.2013.10.012

49. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. New England Journal of Medicine. 2015;373: 895–907. doi:10.1056/NEJMoa1502214

50. Osório J. Gene regulation: Landscape and mechanisms of transcription factor cooperativity. Nat Rev Genet. 2016;17: 5–5. doi:10.1038/nrg.2015.11

51. Bird A. DNA methylation patterns and epigenetic memory. Genes & development. 2002;16: 6–21. doi:10.1101/gad.947102

52. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet. 2012;13: 97–109. doi:10.1038/nrg3142

53. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33: 245–254. doi:10.1038/ng1089

54. Bird A. Perceptions of epigenetics. Nature. 2007;447: 396–398. doi:10.1038/nature05913

55. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128: 693–705. doi:10.1016/j.cell.2007.02.005

56. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. Nat Rev Genet. 2010;11: 191–203. doi:10.1038/nrg2732

57. Fouse SD, Nagarajan RP, Costello JF. Genome-scale DNA methylation analysis. Epigenomics. 2010;2: 105–117. doi:10.2217/epi.09.35

58. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13: 484–492. doi:10.1038/nrg3230

59. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500: 477–481. doi:10.1038/nature12433

60. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11: 204–220. doi:10.1038/nrg2719

61. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502: 472–479. doi:10.1038/nature12750

62. Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol. 2010;11: 607–620. doi:10.1038/nrm2950

63. Sado T, Okano M, Li E, Sasaki H. De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. Development. 2004;131: 975–982. doi:10.1242/dev.00995

64. Kelly TK, Miranda TB, Liang G, Berman BP, Lin JC, Tanay A, et al. H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. Mol Cell. 2010;39: 901–911. doi:10.1016/j.molcel.2010.08.026

65. Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, et al. Frequent switching of Polycomb repressive marks and DNA hypermethylation in

the PC3 prostate cancer cell line. Proc Natl Acad Sci USA. 2008;105: 12979–12984. doi:10.1073/pnas.0806437105

66. Wajed SA, Laird PW, DeMeester TR. DNA Methylation: An Alternative Pathway to Cancer. Ann Surg. 2001;234: 10–20.

67. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene Body Methylation can alter Gene Expression and is a Therapeutic Target in Cancer. Cancer Cell. 2014;26: 577–590. doi:10.1016/j.ccr.2014.07.028

68. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, et al. Dynamic changes in the human methylome during differentiation. Genome Res. 2010;20: 320–331. doi:10.1101/gr.101907.109

69. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. eLife. 2013;2: e00523. doi:10.7554/eLife.00523

70. Ling C, Del Guerra S, Lupi R, Rönn T, Granhall C, Luthman H, et al. Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia. 2008;51: 615–622. doi:10.1007/s00125-007-0916-5

71. Dayeh T, Volkov P, Salö S, Hall E, Nilsson E, Olsson AH, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. PLoS genetics. 2014;10: e1004160. doi:10.1371/journal.pgen.1004160

72. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou Y-H, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. Hum Mol Genet. 2015;24: 4464–4479. doi:10.1093/hmg/ddv161

73. Nilsson E, Matte A, Perfilyev A, de Mello VD, Käkelä P, Pihlajamäki J, et al. Epigenetic Alterations in Human Liver From Subjects With Type 2 Diabetes in Parallel With Reduced Folate Levels. J Clin Endocrinol Metab. 2015;100: E1491–E1501. doi:10.1210/jc.2015-3204

74. Nitert MD, Dayeh T, Volkov P, Elgzyri T, Hall E, Nilsson E, et al. Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. Diabetes. 2012;61: 3322–3332. doi:10.2337/db11-1653

75. Vaag A, Brøns C, Gillberg L, Hansen NS, Hjort L, Arora GP, et al. Genetic, nongenetic and epigenetic risk determinants in developmental programming of type 2 diabetes. Acta Obstet Gynecol Scand. 2014;93: 1099–1108. doi:10.1111/aogs.12494

76. Taudt A, Colomé-Tatché M, Johannes F. Genetic sources of population epigenomic variation. Nat Rev Genet. 2016;17: 319–332. doi:10.1038/nrg.2016.45

77. McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, et al. Identification of genetic variants that affect histone modifications in human cells. Science. 2013;342: 747–749. doi:10.1126/science.1242429

78. Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, et al. Extensive variation in chromatin states across humans. Science. 2013;342: 750–752. doi:10.1126/science.1242510

79. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat Neurosci. 2016;19: 48–54. doi:10.1038/nn.4182

80. Olsson AH, Volkov P, Bacos K, Dayeh T, Hall E, Nilsson EA, et al. Genome-Wide Associations between Genetic and Epigenetic Variation Influence mRNA Expression and Insulin Secretion in Human Pancreatic Islets. PLoS Genet. 2014;10: e1004735. doi:10.1371/journal.pgen.1004735

81. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biology. 2014;15: R73. doi:10.1186/gb-2014-15-5-r73

82. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA. 1992;89: 1827–1831.

83. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium® assay. Epigenomics. 2009;1: 177–200. doi:10.2217/epi.09.14

84. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98: 288–295. doi:10.1016/j.ygeno.2011.07.007

85. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11: 587. doi:10.1186/1471-2105-11-587

86. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011;3: 771–784. doi:10.2217/epi.11.105

87. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium

HumanMethylation450 microarray. Epigenetics. 2013;8: 203–209. doi:10.4161/epi.23470

88. Price EM, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics & Chromatin. 2013;6: 4. doi:10.1186/1756-8935-6-4

89. Multhaup ML, Seldin MM, Jaffe AE, Lei X, Kirchner H, Mondal P, et al. Mouse-Human Experimental Epigenetic Analysis Unmasks Dietary Targets and Genetic Liability for Diabetic Phenotypes. Cell Metabolism. 2015;21: 138–149. doi:10.1016/j.cmet.2014.12.014

90. Morris TJ, Beck S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. Methods. 2015;72: 3–8. doi:10.1016/j.ymeth.2014.08.011

91. Veillard A-C, Datlinger P, Laczik M, Squazzo S, Bock C. Diagenode® Premium RRBS technology: cost-effective DNA methylation mapping with superior coverage. Nat Meth. 2016;13. doi:10.1038/nmeth.f.391

92. Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, et al. A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip(®) for methylome profiling. PLoS ONE. 2012;7: e50233. doi:10.1371/journal.pone.0050233

93. Yong W-S, Hsu F-M, Chen P-Y. Profiling genome-wide DNA methylation. Epigenetics Chromatin. 2016;9: 26. doi:10.1186/s13072-016-0075-3

94. Bock C. Analysing and interpreting DNA methylation data. Nat Rev Genet. 2012;13: 705–719. doi:10.1038/nrg3273

95. Brøns C, Jensen CB, Storgaard H, Alibegovic A, Jacobsen S, Nilsson E, et al. Mitochondrial function in skeletal muscle is normal and unrelated to insulin action in young men born with low birth weight. The Journal of Clinical Endocrinology and Metabolism. 2008;93: 3885–3892. doi:10.1210/jc.2008-0630

96. Jørgensen SW, Brøns C, Bluck L, Hjort L, Færch K, Thankamony A, et al. Metabolic response to 36 hours of fasting in young men born small vs appropriate for gestational age. Diabetologia. 2014; doi:10.1007/s00125-014-3406-6

97. Keltanen T, Sajantila A, Valonen T, Vanhala T, Lindroos K. Measuring postmortem glycated hemoglobin - A comparison of three methods. Leg Med (Tokyo). 2013;15: 72–78. doi:10.1016/j.legalmed.2012.09.002

98. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from

fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28: 412–419.

99. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007;81: 559–575. doi:10.1086/519795

100. Schwender H, Li Q, Neumann C, Taub M, Ruczinski I. trio: Testing of SNPs and SNP Interactions in Case-Parent Trio Studies. R package version 3.0.0. 2013;

101. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19: 185–193.

102. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29: 189–196. doi:10.1093/bioinformatics/bts680

103. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8: 118–127. doi:10.1093/biostatistics/kxj037

104. Fuller CW, Middendorf LR, Benner SA, Church GM, Harris T, Huang X, et al. The challenges of sequencing by synthesis. Nat Biotechnol. 2009;27: 1013–1023. doi:10.1038/nbt.1585

105. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17: 10–12.

106. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27: 1571–1572. doi:10.1093/bioinformatics/btr167

107. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359. doi:10.1038/nmeth.1923

108. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol. 2012;13: R83. doi:10.1186/gb-2012-13-10-r83

109. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4: 249–264. doi:10.1093/biostatistics/4.2.249

110. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12: 323. doi:10.1186/1471-2105-12-323

111. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013;

112. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28: 1353–1358. doi:10.1093/bioinformatics/bts163

113. Benjamini Y, Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. 1995;57: 289–300.

114. The Python Language Reference [Internet]. [cited 21 Jul 2016]. Available: https://docs.python.org/3/reference/

115. McCarthy MI. Genomics, Type 2 Diabetes, and Obesity. New England Journal of Medicine. 2010;363: 2339–2350. doi:10.1056/NEJMra0906948

116. Singhal J, Nagaprashantha L, Vatsyayan R, Awasthi S, Singhal SS. RLIP76, a Glutathione-Conjugate Transporter, Plays a Major Role in the Pathogenesis of Metabolic Syndrome. PLOS ONE. 2011;6: e24688. doi:10.1371/journal.pone.0024688

117. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012;44: 659–669. doi:10.1038/ng.2274

118. Consortium GLG. Discovery and refinement of loci associated with lipid levels. Nature Genetics. 2013;45: 1274–1283. doi:10.1038/ng.2797

119. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42: 105–116. doi:10.1038/ng.520

120. Williams MJ, Almén MS, Fredriksson R, Schiöth HB. What model organisms and interactomics can reveal about the genetics of human obesity. Cell Mol Life Sci. 2012;69: 3819–3834. doi:10.1007/s00018-012-1022-5

121. Mahajan A, Sim X, Ng HJ, Manning A, Rivas MA, Highland HM, et al. Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. PLoS Genet. 2015;11: e1004876. doi:10.1371/journal.pgen.1004876

122. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. BMC Genet. 2009;10: 23. doi:10.1186/1471-2156-10-23

123. Stoffel M, Stein R, Wright CV, Espinosa R, Le Beau MM, Bell GI. Localization of human homeodomain transcription factor insulin promoter factor 1 (IPF1) to chromosome band 13q12.1. Genomics. 1995;28: 125–126.

124. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22: 1760–1774. doi:10.1101/gr.135350.111

125. Long HK, King HW, Patient RK, Odom DT, Klose RJ. Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. Nucleic Acids Res. 2016; doi:10.1093/nar/gkw258

126. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet. 2010;42: 949–960. doi:10.1038/ng.685

127. Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, et al. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. Age (Dordr). 2014;36. doi:10.1007/s11357-014-9648-x

128. Bacos K, Gillberg L, Volkov P, Olsson AH, Hansen T, Pedersen O, et al. Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. Nature Communications. 2016;7. doi:10.1038/ncomms11089

129. Pociot F, McDermott MF. Genetics of type 1 diabetes mellitus. Genes and immunity. 2002;3: 235–249. doi:10.1038/sj.gene.6363875

130. Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004. Tissue antigens. 2004;64: 631–649. doi:10.1111/j.1399-0039.2004.00327.x

131. Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI, et al. FTO genotype is associated with phenotypic variability of body mass index. Nature. 2012;490: 267–272. doi:10.1038/nature11401

132. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al. Common variants at 10 genomic loci influence hemoglobin A□(C) levels via glycemic and nonglycemic pathways. Diabetes. 2010;59: 3229–3239. doi:10.2337/db10-0502

133. Volkmar M, Dedeurwaerder S, Cunha DA, Ndlovu MN, Defrance M, Deplus R, et al. DNA methylation profiling identifies epigenetic dysregulation in pancreatic

islets from type 2 diabetic patients. EMBO J. 2012;31: 1405–1426. doi:10.1038/emboj.2011.503

134. Ziller MJ, Hansen KD, Meissner A, Aryee MJ. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nat Meth. 2015;12: 230–232. doi:10.1038/nmeth.3152

135. Xiao Y, Camarillo C, Ping Y, Arana TB, Zhao H, Thompson PM, et al. The DNA Methylome and Transcriptome of Different Brain Regions in Schizophrenia and Bipolar Disorder. PLoS One. 2014;9. doi:10.1371/journal.pone.0095875

136. Bell RE, Golan T, Malcov H, Amar D, Salamon A, Liron T, et al. Enhancer methylation dynamics contribute to cancer plasticity and patient mortality. Genome Res. 2016; gr.197194.115. doi:10.1101/gr.197194.115

137. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman I, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet. 2014;46: 136–143. doi:10.1038/ng.2870

138. Kaneto H, Miyatsuka T, Kawamori D, Yamamoto K, Kato K, Shiraiwa T, et al. PDX-1 and MafA play a crucial role in pancreatic beta-cell differentiation and maintenance of mature beta-cell function. Endocr J. 2008;55: 235–252.

139. Ahlgren U, Jonsson J, Jonsson L, Simu K, Edlund H. β-Cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the β-cell phenotype and maturity onset diabetes. Genes Dev. 1998;12: 1763–1768.

140. Yang BT, Dayeh TA, Volkov PA, Kirkpatrick CL, Malmgren S, Jing X, et al. Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. Molecular endocrinology (Baltimore, Md). 2012;26: 1203–1212. doi:10.1210/me.2012-1004

141. Houseman EA, Kelsey KT, Wiencke JK, Marsit CJ. Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective. BMC Bioinformatics. 2015;16: 95. doi:10.1186/s12859-015-0527-y

142. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics. 2014;30: 1431–1439. doi:10.1093/bioinformatics/btu029

Study I

OPEN ACCESS Freely available online

PLOS | GENETICS

A Six Months Exercise Intervention Influences the Genome-wide DNA Methylation Pattern in Human Adipose Tissue

Tina Rönn[1]*, Petr Volkov[1], Cajsa Davegårdh[1], Tasnim Dayeh[1], Elin Hall[1], Anders H. Olsson[1], Emma Nilsson[1], Åsa Tornberg[2], Marloes Dekker Nitert[3], Karl-Fredrik Eriksson[4], Helena A. Jones[5], Leif Groop[6], Charlotte Ling[1]*

1 Department of Clinical Sciences, Epigenetics and Diabetes, Lund University Diabetes Centre, CRC, Malmö, Sweden, **2** Department of Health Sciences, Division of Physiotherapy, Lund University, Lund, Sweden, **3** School of Medicine, Royal Brisbane Clinical School, The University of Queensland, Herston, Queensland, Australia, **4** Department of Clinical Sciences, Vascular Diseases, Lund University, Malmö, Sweden, **5** Department of Experimental Medical Science, Division of Diabetes, Metabolism and Endocrinology, Lund University, BMC C11, Lund, Sweden, **6** Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, CRC, Malmö, Sweden

Abstract

Epigenetic mechanisms are implicated in gene regulation and the development of different diseases. The epigenome differs between cell types and has until now only been characterized for a few human tissues. Environmental factors potentially alter the epigenome. Here we describe the genome-wide pattern of DNA methylation in human adipose tissue from 23 healthy men, with a previous low level of physical activity, before and after a six months exercise intervention. We also investigate the differences in adipose tissue DNA methylation between 31 individuals with or without a family history of type 2 diabetes. DNA methylation was analyzed using Infinium HumanMethylation450 BeadChip, an array containing 485,577 probes covering 99% RefSeq genes. Global DNA methylation changed and 17,975 individual CpG sites in 7,663 unique genes showed altered levels of DNA methylation after the exercise intervention ($q<0.05$). Differential mRNA expression was present in 1/3 of gene regions with altered DNA methylation, including *RALBP1*, *HDAC4* and *NCOR2* ($q<0.05$). Using a luciferase assay, we could show that increased DNA methylation *in vitro* of the *RALBP1* promoter suppressed the transcriptional activity ($p=0.03$). Moreover, 18 obesity and 21 type 2 diabetes candidate genes had CpG sites with differences in adipose tissue DNA methylation in response to exercise ($q<0.05$), including *TCF7L2* (6 CpG sites) and *KCNQ1* (10 CpG sites). A simultaneous change in mRNA expression was seen for 6 of those genes. To understand if genes that exhibit differential DNA methylation and mRNA expression in human adipose tissue *in vivo* affect adipocyte metabolism, we silenced *Hdac4* and *Ncor2* respectively in 3T3-L1 adipocytes, which resulted in increased lipogenesis both in the basal and insulin stimulated state. In conclusion, exercise induces genome-wide changes in DNA methylation in human adipose tissue, potentially affecting adipocyte metabolism.

Citation: Rönn T, Volkov P, Davegårdh C, Dayeh T, Hall E, et al. (2013) A Six Months Exercise Intervention Influences the Genome-wide DNA Methylation Pattern in Human Adipose Tissue. PLoS Genet 9(6): e1003572. doi:10.1371/journal.pgen.1003572

Editor: John M. Greally, Albert Einstein College of Medicine, United States of America

Received January 4, 2013; **Accepted** May 2, 2013; **Published** June 27, 2013

Funding: This work was supported by grants from the Swedish Research Council (CL and LG) and Lund University Diabetes Centre (LUDC), the Knut & Alice Wallenbergs stiftelse, Fredrik & Ingrid Thurings stiftelse (TR), Kungliga Fysiografiska sällskapet (TR), Tore Nilssons stiftelse (TR), Påhlssons stiftelse (CL), Novonordisk foundation (CL), ALF (CL), Diabetes förbundet (CL), Söderbergs stiftelse (CL) and by an EU grant (ENGAGE; LG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tina.ronn@med.lu.se (TR); charlotte.ling@med.lu.se (CL)

Introduction

A sedentary lifestyle, a poor diet and new technologies that reduce physical activity cause health problems worldwide, as reduced energy expenditure together with increased energy intake lead to weight gain and increased cardiometabolic health risks [1]. Obesity is an important predictor for the development of both type 2 diabetes (T2D) and cardiovascular diseases, which suggests a central role for adipose tissue in the development of these conditions [2]. Adipose tissue is an endocrine organ affecting many metabolic pathways, contributing to total glucose homeostasis [2]. T2D is caused by a complex interplay of genetic and lifestyle factors [3], and a family history of T2D has been associated with reduced physical fitness and an increased risk of the disease [4–6]. Individuals with high risk of developing T2D strongly benefit from non-pharmacological interventions, involving diet and exercise [7,8]. Exercise is important for physical health, including weight maintenance and its beneficial effects on triglycerides, cholesterol and blood pressure, suggestively by activating a complex program of transcriptional changes in target tissues.

Epigenetic mechanisms such as DNA methylation are considered to be important in phenotype transmission and the development of different diseases [9]. The epigenetic pattern is mainly established early in life and thereafter maintained in differentiated cells, but age-dependent alterations still have the potential to modulate gene expression and translate environmental factors into phenotypic traits [10–13]. In differentiated mamma-

Author Summary

Given the important role of epigenetics in gene regulation and disease development, we here present the genome-wide DNA methylation pattern of 476,753 CpG sites in adipose tissue obtained from healthy men. Since environmental factors potentially change metabolism through epigenetic modifications, we examined if a six months exercise intervention alters the DNA methylation pattern as well as gene expression in human adipose tissue. Our results show that global DNA methylation changes and 17,975 individual CpG sites alter the levels of DNA methylation in response to exercise. We also found differential DNA methylation of 39 candidate genes for obesity and type 2 diabetes in human adipose tissue after exercise. Additionally, we provide functional proof that genes, which exhibit both differential DNA methylation and gene expression in human adipose tissue in response to exercise, influence adipocyte metabolism. Together, this study provides the first detailed map of the genome-wide DNA methylation pattern in human adipose tissue and links exercise to altered adipose tissue DNA methylation, potentially affecting adipocyte metabolism.

lian cells, DNA methylation usually occurs in the context of CG dinucleotides (CpGs) and is associated with gene repression [14]. Changes in epigenetic profiles are more common than genetic mutations and may occur in response to environmental, behavioural, psychological and pathological stimuli [15]. Furthermore, genetic variation not associated with a phenotype could nonetheless affect the extent of variability of that phenotype through epigenetic mechanisms, such as DNA methylation. It is not known whether epigenetic modifications contribute to the cause or transmission of T2D between generations. Recent studies in human skeletal muscle and pancreatic islets point towards the involvement of epigenetic modifications in the regulation of genes important for glucose metabolism and the pathogenesis of T2D [11,12,16–21]. However, there is limited information about the regulation of the epigenome in human adipose tissue [22].

The mechanisms behind the long-lasting effects of regular exercise are not fully understood, and most studies have focused on cellular and molecular changes in skeletal muscle. Recently, a global study of DNA methylation in human skeletal muscle showed changes in the epigenetic pattern in response to long-term exercise [23]. The aims of this study were to: 1) explore genome-wide levels of DNA methylation before and after a six months exercise intervention in adipose tissue from healthy, but previously sedentary men; 2) investigate the differences in adipose tissue DNA methylation between individuals with or without a family history of T2D; 3) relate changes in DNA methylation to adipose tissue mRNA expression and metabolic phenotypes *in vitro*.

Results

Baseline characteristics of individuals with (FH^+) or without (FH^-) a family history of type 2 diabetes

A total of 31 men, 15 FH^+ and 16 FH^-, had subcutaneous adipose tissue biopsies taken at baseline. The FH^+ and FH^- individuals were group-wise matched for age, gender, BMI and VO_{2max} at inclusion, and there were no significant differences between FH^+ and FH^- individuals, respectively (Table S1). DNA methylation in the adipose tissue was analyzed using the Infinium HumanMethylation450 BeadChip array. After quality control (QC), DNA methylation data was obtained for a total number of 476,753 sites. No individual CpG site showed a significant difference in DNA methylation between FH^+ and FH^- men after false discovery rate (FDR) correction ($q>0.05$) [24]. Additionally, there were no global differences between the FH^+ and FH^- individuals when calculating the average DNA methylation based on genomic regions (Figure 1a) or CpG content (Figure 1b; $q>0.05$).

Clinical outcome and global changes in adipose tissue DNA methylation in response to exercise

Subcutaneous adipose tissue biopsies were taken from 23 men both before and after exercise, followed by successful DNA extraction and analysis of DNA methylation using the Infinium HumanMethylation450 BeadChip array. Since we found no significant differences in DNA methylation between FH^+ and FH^- men at baseline, the two groups were combined when examining the impact of exercise on DNA methylation in adipose tissue. In Table 1 the clinical and metabolic outcomes of the exercise intervention are presented for these 23 men, showing a significant decrease in waist circumference, waist/hip ratio, diastolic blood pressure, and resting heart rate, whereas a significant increase was seen for VO_{2max} and HDL.

To evaluate the global human methylome in adipose tissue, we first calculated the average level of DNA methylation in groups based on either the functional genome distribution (Figure 1a), or the CpG content and neighbourhood context (Figure 1b). We also present the average level of DNA methylation separately for the Infinium I ($n = 126{,}804$) and Infinium II ($n = 326{,}640$) assays due to different β-value distributions for these assays [25]. When evaluating Infinium I assays in relation to nearest gene, the global level of DNA methylation after exercise increased in the 3′ untranslated region (UTR; $q<0.05$), whereas a decrease was seen in the region 1500–200 bp upstream of transcription start (TSS1500), TSS200, 5′UTR and within the first exon (1st Exon; $q<0.05$). The global DNA methylation level of Infinium II assays increased significantly ($q<0.05$) after exercise within all regions except TSS200 (Figure 1c and Table S2). In general, the average level of DNA methylation was low in the region from TSS1500 to the 1st Exon (5–36%), whereas the gene body, the 3′UTR and intergenic region displayed average DNA methylation levels ranging from 43–72% (Figure 1c and Table S2). When evaluating global DNA methylation based on CpG content and distance to CpG islands, average DNA methylation for Infinium I assays decreased significantly after exercise in CpG islands, whereas an increase was seen in northern and southern shelves (regions 2000–4000 bp distant from CpG islands) as well as in the open sea (regions further away from a CpG island) ($q<0.05$; Figure 1d and Table S2). For Infinium II assays, average DNA methylation was significantly increased in all regions after the exercise intervention ($q<0.05$; Figure 1d and Table S2). The global level of DNA methylation was low within CpG islands (9–21%), intermediate within the shores (2000 bp regions flanking the CpG islands; 31–44%), whereas the shelves and the open sea showed the highest level of DNA methylation (67–76%; Figure 1d and Table S2). Although technical variation between probe types has been reported for the Infinium HumanMethylation450 BeadChip array, seen as a divergence between the β-values distribution retrieved from the Infinium I and II assays [25], the global differences in DNA methylation we observe between probe types are more likely a result of skewed GC content due to the design criteria of the two different assays. Infinium I assays have significantly more CpGs within the probe body than the Infinium II assays, and 57% are annotated to CpG islands, whereas most

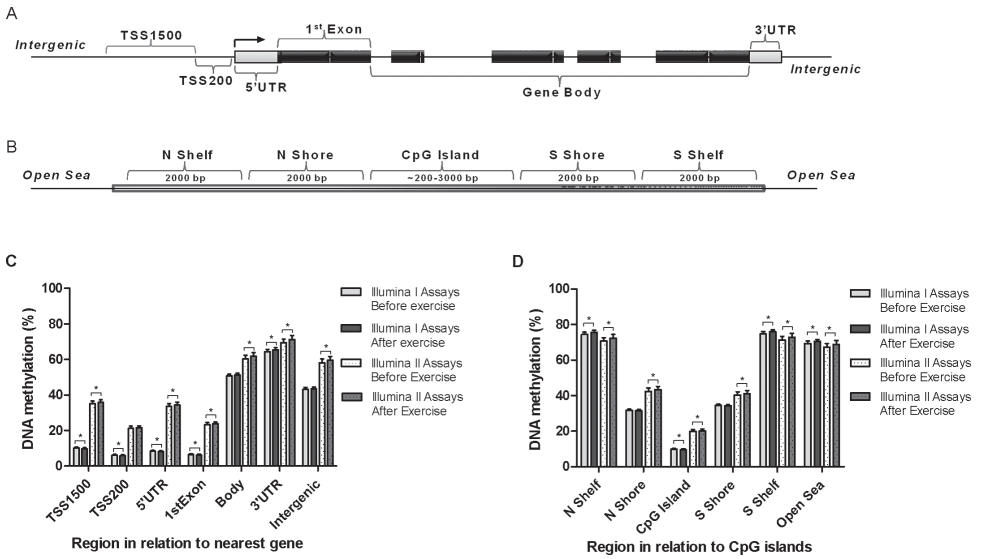


Figure 1. Location of analyzed CpG sites and global DNA methylation in human adipose tissue. All CpG sites analyzed on the Infinium HumanMethylation450 BeadChip are mapped to gene regions based on functional genome distribution (A) and to CpG island regions based on CpG content and neighbourhood context (B). In the lower panels, global DNA methylation in human adipose tissue is shown for each gene region (C) and for CpG island regions (D). Global DNA methylation is calculated as average DNA methylation based on all CpG sites in each region on the chip, and presented separately for Infinium I and Infinium II assays, respectively. Data is presented as mean ± SD. TSS, proximal promoter, defined as 200 bp (basepairs) or 1500 bp upstream of the transcription start site; UTR, untranslated region; CpG island, 200 bp (or more) stretch of DNA with a C+G content of 50% and an observed CpG/expected CpG in excess of 0.6; Shelf, regions flanking island shores, i.e., covering 2000–4000 bp distant from the CpG island; Shore: the flanking region of CpG islands, 0–2000 bp. *Significant difference between average DNA methylation before versus after exercise, $q<0.05$.
doi:10.1371/journal.pgen.1003572.g001

Infinium II assays have less than three underlying CpGs in the probe and only 21% are designated as CpG islands [26].

DNA methylation of individual CpG sites in human adipose tissue is influenced by exercise

We next investigated if there was a difference in DNA methylation in any of the 476,753 analyzed individual CpG sites in adipose tissue in response to exercise. A flowchart of the analysis process is found in Figure 2. SNPs within the probe were not a criterion for exclusion in this analysis, as the participants are their own controls, thereby excluding genetic variation within the tested pairs. Applying FDR correction ($q<0.05$) resulted in 17,975 CpG sites, corresponding to 7,663 unique genes, that exhibit differential DNA methylation in adipose tissue after exercise. Among these 17,975 individual sites, 16,470 increased and 1,505 decreased the level of DNA methylation in response to exercise, with absolute changes in DNA methylation ranging from 0.2–10.9% (Figure 3a–b). Aiming for biological relevance, we further filtered our results requiring the average change in DNA methylation (β-value) for each CpG site to be ≥5% before vs. after exercise. Adding the criteria with a ≥5% change in DNA methylation resulted in 1,009 significant individual CpG sites: 911 with increased and 98 with decreased levels of DNA methylation in response to the six months exercise intervention. Of those, 723 sites are annotated to one or more genes, and correspond to 641 unique gene IDs. A comparison of our 1,009 significant CpG sites with Infinium probes reported to cross-react to alternative genomic locations [27] showed only one probe with 50 bases and 14 probes with 49 bases matching to an alternative genomic location. Data of the most significant CpG sites ($q<0.005$) and the sites that exhibit the greatest change in adipose tissue DNA methylation (difference in DNA methylation >8%) in response to exercise are presented in Table 2–3 and included *ITPR2* and *TSTD1* for increased, and *LTBP4* for decreased DNA methylation. We found 7 CpG sites in this list to be targeted by Infinium probes reported to cross-react to alternative genomic locations (47 or 48 bases) [27]. Additionally, to investigate the possibility that the changes we see in response to exercise is rather an effect of epigenetic drift over time, we compared our 1,009 differentially methylated CpG sites ($q<0.05$, difference in β-value>5%) with three studies reporting aging-differentially methylated regions (a-DMRs) in a total of 597 unique positions [28–30]. Secondly we tested for association between age and the level of DNA methylation in the 31 individuals included at baseline in this study, representing a more valid age range (30–45 years) and tissue for the current hypothesis. We found no overlap between previously published a-DMRs or the age-associated CpG sites within our study (18 CpG sites; $p<1\times10^{-5}$), and the CpG sites differentially methylated after the exercise intervention.

The genomic distribution of individual CpG sites with a significant change in DNA methylation ≥5% with exercise is shown in Figure 3c–d, in comparison to all probes located on the Infinium HumanMethylation450 BeadChip and passing QC. The distribution is based on location in relation to the functional genome distribution (Figure 3c) or CpG content and distance to CpG islands

Table 1. Clinical characteristics of study participants ($n = 23$) with DNA methylation data both before (baseline) and after the exercise intervention.

Characteristics	Baseline	After exercise	*p*-value
Age (years)	37.3±4.4	-	
Weight (kg)	91.8±11.0	90.8±11.6	0.18
BMI (kg/m^2)	28.2±2.9	27.9±3.1	0.18
Waist circumference (cm)	97.7±8.6	95.7±8.7	0.02
Waist/hip ratio	0.93±0.05	0.92±0.06	0.01
Fatmass (%)	22.8±6.0	23.1±6.6	0.59
Fasting glucose (mmol/L)	5.01±0.64	4.95±0.59	0.51
2 h OGTT glucose (mmol/L)	6.17±1.02	5.86±1.47	0.32
HbA1c (%)	4.31±0.31	4.31±0.34	1.00
Fasting insulin (μU/mL)	6.60±2.41	6.80±2.86	0.63
VO_{2max} (mL/kg/min)	33.1±4.6	36.2±6.2	0.003
Systolic BP (mmHg)	132.5±10.2	129.9±11.8	0.34
Diastolic BP (mmHg)	79.3±9.3	74.8±10.7	0.04
Pulse (beats/min)	73.9±10.6	67.3±11.2	0.03
Total cholesterol (mmol/L)	4.99±0.71	4.63±1.12	0.07
Triglycerides (mmol/L)	1.63±1.30	1.26±0.98	0.20
LDL (mmol/L)	3.36±0.63	3.24±0.63	0.41
HDL (mmol/L)	1.04±0.21	1.11±0.21	0.02
LDL/HDL	3.31±0.89	3.02±0.92	0.053

Data are expressed as mean ± SD, based on paired t-tests and two-tailed *p*-values. BP, blood pressure; LDL, low density lipoprotein; HDL, high density lipoprotein.

doi:10.1371/journal.pgen.1003572.t001

(Figure 3d). We found that the CpG sites with altered level of DNA methylation in response to exercise were enriched within the gene body and in intergenic regions, while the proximal promoter, in particular TSS200 and the 1st exon, had a low proportion of differentially methylated CpG sites ($p = 7\times10^{-20}$; Figure 3c). In relation to CpG content and distance to CpG islands, the region with the highest proportion of significant CpG sites compared to the distribution on the array was in the open sea, *i.e.*, regions more distant from a CpG island than 4000 bp. In contrast, the number of significant CpG sites found within the CpG islands was only half of what would be expected ($p = 2\times10^{-31}$; Figure 3d).

Exercise induces overlapping changes in DNA methylation and mRNA expression

An increased level of DNA methylation has previously been associated with transcription repression [14]. We therefore related changes in adipose tissue DNA methylation of individual CpG-sites ($q<0.05$ and difference in mean β-values ≥5%) with changes in mRNA expression of the same gene ($q<0.05$) in response to exercise (Figure 2). We identified 236 CpG sites in 197 individual gene regions that exhibit differential DNA methylation together with a significant change in adipose tissue mRNA expression of the corresponding gene after exercise. Of these, 143 CpG sites (61%) connected to 115 genes showed an inverse relation to mRNA expression. After exercise, 139 CpG sites showed an increase in DNA methylation and a corresponding decrease in mRNA expression, including a gene for one of the GABA receptors (*GABBR1*), several genes encoding histone modifying enzymes (*EHMT1*, *EHMT2* and *HDAC4*) and a transcriptional co-repressor (*NCOR2*). Only four CpG sites were found to decrease in the level of DNA methylation with a concomitant increase in mRNA expression. Table S3 shows all significant results of DNA methylation sites with an inverse relation to mRNA expression in human adipose tissue before vs. after exercise.

DNA methylation *in vitro* decreases reporter gene expression

RALBP1 belongs to the genes that exhibit increased DNA methylation in the promoter region in parallel with decreased mRNA expression in adipose tissue in response to exercise (Figure 4a–b and Table S3). It has previously been shown to play a central role in the pathogenesis of metabolic syndrome [31] and to be involved in insulin-stimulated Glut4 trafficking [32]. We proceeded to functionally test if increased DNA methylation of the promoter of *RALBP1* may cause decreased gene expression using a reporter gene construct in which 1500 bp of DNA of the human *RALBP1* promoter was inserted into a luciferase expression plasmid that completely lacks CpG dinucleotides. The reporter construct could thereby be used to study the effect of promoter DNA methylation on the transcriptional activity. The construct was methylated using two different methyltransferases; SssI and HhaI, which methylate all CpG sites or only the internal cytosine residue in a GCGC sequence, respectively.

Increased DNA methylation of the *RALBP1* promoter, as measured by luciferase activity, suppressed the transcriptional activity of the promoter ($p = 0.028$, Figure 4c). When the *RALBP1* reporter construct was methylated *in vitro* using SssI (CG, 94 CpG sites), the transcriptional activity was almost completely disrupted (1.4±0.5), whereas the HhaI enzyme (GCGC, methylating 14 CpG sites) suppressed the transcriptional activity to a lesser extent (23.4±11.6), compared with the transcriptional activity of the mock-methylated control construct (448.2±201.7; Figure 4c).

DNA methylation of obesity and type 2 diabetes candidate genes in human adipose tissue

We proceeded to investigate if candidate genes for obesity or T2D, identified using genome-wide association studies [3], are found among the genes exhibiting changed levels of DNA methylation in adipose tissue in response to six months exercise. Among all 476,753 CpG sites analyzed on the Infinium HumanMethylation450 BeadChip and passing QC, 1,351 sites mapped to 53 genes suggested to contribute to obesity in the review by McCarthy, and 1,315 sites mapped to 39 genes suggested to contribute to T2D [3]. We found 24 CpG sites located within 18 of the candidate genes for obesity with a difference in DNA methylation in adipose tissue in response to the exercise intervention ($q<0.05$, Table 4). Additionally, two of those genes (*CPEB4* and *SDCCAG8*) showed concurrent inverse change in mRNA expression after exercise ($q<0.05$). Among the T2D candidate genes, 45 CpG sites in 21 different genes were differentially methylated ($q<0.05$) in adipose tissue before vs. after exercise (Table 5). Of note, 10 of these CpG sites mapped to *KCNQ1* and 6 sites mapped to *TCF7L2*. A simultaneous change in mRNA expression was seen for four of the T2D candidate genes (*HHEX*, *IGF2BP2*, *JAZF1* and *TCF7L2*) where mRNA expression decreased while DNA methylation increased in response to exercise ($q<0.05$, Table 5).

Silencing of *Hdac4* and *Ncor2* in 3T3-L1 adipocytes is associated with increased lipogenesis

To further understand if the genes that exhibit differential DNA methylation and mRNA expression in adipose tissue *in vivo* affect adipocyte metabolism, we silenced the expression of selected genes in 3T3-L1 adipocytes using siRNA and studied its effect on lipogenesis. Two of the genes where we found increased DNA

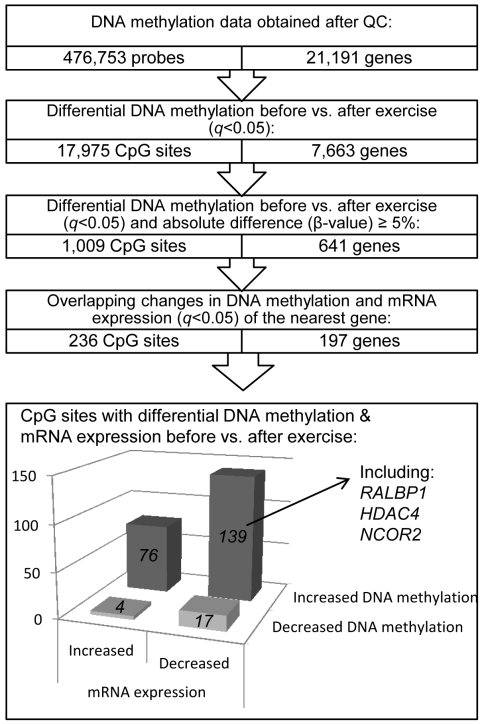


Figure 2. Analysis flowchart.
doi:10.1371/journal.pgen.1003572.g002

methylation in parallel with decreased mRNA expression in human adipose tissue in response to exercise (Figure 5a–d and Table S3) were selected for functional studies in a 3T3-L1 adipocyte cell line. *HDAC4* was further a strong candidate due to multiple affected CpG sites within the gene, and both *HDAC4* and *NCOR2* are biologically interesting candidates in adipose tissue and the pathogenesis of obesity and type 2 diabetes [33–35]. Silencing of *Hdac4* and *Ncor2* in the 3T3-L1 adipocytes resulted in 74% reduction in the Hdac4 protein level (1.00±0.50 vs. 0.26±0.20, $p = 0.043$; Figure 5e) while the *Ncor2* mRNA level was reduced by 56% (1.00±0.19 vs. 0.44±0.08, $p = 0.043$; Figure 5f) of control after transfection with siRNA for 72 hours and 24 h, respectively. Lipogenesis was nominally increased in the basal state (1.00±0.26 vs. 1.44±0.42, $p = 0.079$) and significantly increased in response to 0.1 nM insulin (1.16±0.30 vs. 1.52±0.34, $p = 0.043$) in 3T3-L1 adipocytes with decreased Hdac4 levels (Figure 5g). Decreased Ncor2 levels also resulted in increased lipogenesis in the basal (1.00±0.19 vs. 1.19±0.19, $p = 0.043$) and insulin stimulated (1 nM; 1.38±0.17 vs. 1.73±0.32, $p = 0.043$) state (Figure 5h).

Technical validation of Infinium HumanMethylation450 BeadChip DNA methylation data

To technically validate the DNA methylation data from the Infinium HumanMethylation450 BeadChips, we compared the genome-wide DNA methylation data from one adipose tissue sample analyzed at four different occasions. Technical reproducibility was observed between all samples, with Pearson's correlation coefficients >0.99 ($p<2.2\times10^{-16}$, Figure S1a). Secondly, we re-analyzed DNA methylation of four CpG sites using Pyrosequencing (PyroMark Q96ID, Qiagen) in adipose tissue of all 23 men both before and after exercise (Table S4). We observed a significant correlation between the two methods for each CpG site ($p<0.05$; Figure S1b), and combining all data points gives a correlation factor of 0.77 between the two methods ($p<0.0001$; Figure S1c).

Discussion

This study highlights the dynamic feature of DNA methylation, described using a genome-wide analysis in human adipose tissue before and after exercise. We show a general global increase in adipose tissue DNA methylation in response to 6 months exercise, but also changes on the level of individual CpG sites, with significant absolute differences ranging from 0.2–10.9%. This data, generated using human adipose tissue biopsies, demonstrate an important role for epigenetic changes in human metabolic processes. Additionally, this study provides a first reference for the DNA methylome in adipose tissue from healthy, middle aged men.

Changes in DNA methylation have been suggested to be a biological mechanism behind the beneficial effects of physical activity [18,36]. In line with this theory, a nominal association between physical activity level and global LINE-1 methylation in leukocytes was recently reported [37]. More important from a metabolic point-of-view, a study investigating the impact of long term exercise intervention on genome-wide DNA methylation in human skeletal muscle was recently published, and showed epigenetic alterations of genes important for T2D pathogenesis and muscle physiology [23]. This relationship between exercise and altered DNA methylation is here expanded to include human adipose tissue, as our data show 17,975 individual CpG sites that exhibit differential DNA methylation in adipose tissue after an exercise intervention, corresponding to 7,663 unique genes throughout the genome. Genome-wide association studies have identified multiple SNPs strongly associated with disease, but still the effect sizes of the common variants influencing for example risk of T2D are modest and in total only explain a small proportion of the predisposition. Importantly, although each variant only contributes with a small risk, these findings have led to improved understanding of the biological basis of disease [3]. Similarly, the absolute changes in DNA methylation observed in response to the exercise intervention are modest, but the large number of affected sites may in combination potentially contribute to a physiological response. Moreover, if the exercise induced differences in DNA methylation is expressed as fold-change instead of absolute differences, we observe changes ranging from 6 to 38%.

In regard to the distribution of analyzed CpG sites, most of the differentially methylated sites were found within the gene bodies and in intergenic regions, and fewer than expected was found in the promoter regions and CpG islands. This is in agreement with previous studies showing that differential DNA methylation is often found in regions other than CpG islands. For example, it was shown that tissue-specific differentially methylated regions in the 5′UTR are strongly underrepresented within CpG islands [38] and that most tissue-specific DNA methylation occurs at CpG island shores rather than the within CpG islands, and also in regions more distant than 2 kb from CpG islands [39]. It has further been proposed that non-CpG island DNA methylation is more dynamic than methylation within CpG islands [40]. The importance of differential DNA methylation within gene bodies is

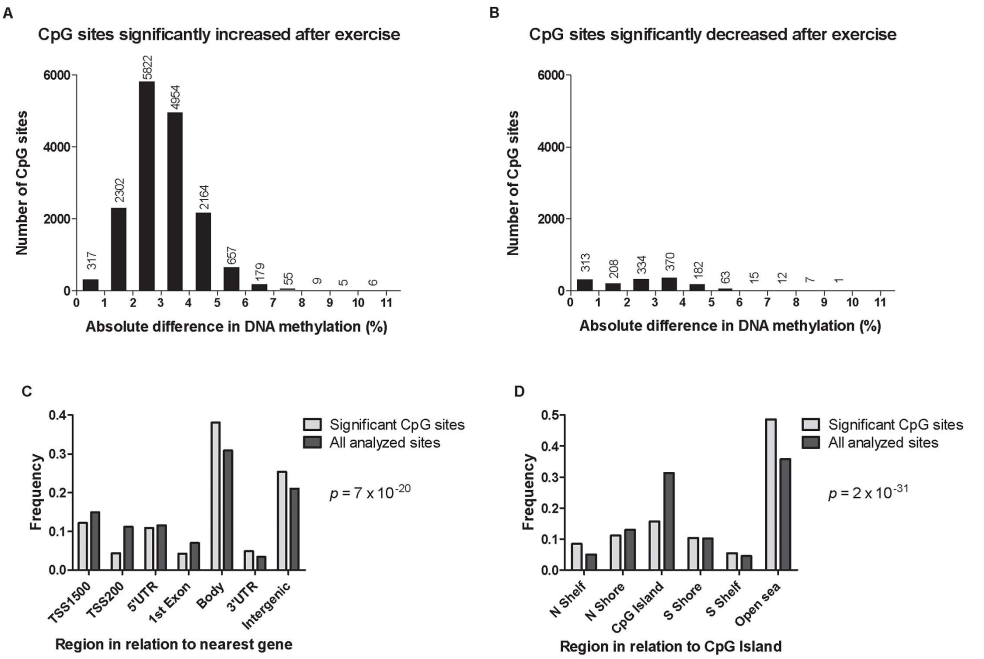


Figure 3. DNA methylation of individual CpG sites. The absolute change in DNA methylation of individual CpG sites with a significant difference after exercise compared with baseline ($q<0.05$) ranges from 0.2–10.9% (A and B). A) Number of sites with increased methylation in adipose tissue in response to exercise ($n = 16,470$). B) Number of sites with decreased DNA methylation in adipose tissue in response to exercise ($n = 1,505$). Panels C and D show the distribution of CpG sites with a significant change ($q<0.05$) and an absolute difference ≥5% in DNA methylation in adipose tissue before versus after exercise, in comparison to all analyzed sites on the Infinium HumanMethylation450 BeadChip. C) Distribution of significant CpG sites vs. all analyzed sites in relation to nearest gene regions. D) Distribution of significant CpG sites vs. all analyzed sites in relation to CpG island regions. *The overall distribution of significant CpG sites compared with all analyzed sites on the Infinium HumanMethylation450 BeadChip was analyzed using a chi^2 test.
doi:10.1371/journal.pgen.1003572.g003

supported by multiple studies showing a positive correlation between gene body methylation and active transcription [40], and that DNA methylation may regulate exon splicing [41,42]. In this study, the exercise intervention associated with a decrease in waist circumference and waist-hip ratio, which suggests reduced abdominal obesity, a phenotype known to be associated with reduced risk of metabolic diseases [43]. Indeed, increased levels of DNA methylation were observed after exercise both in the promoter region and in the gene body of *ITPR2*, a locus previously associated with waist-hip ratio [44]. Furthermore, in addition to increased VO_{2max}, the study participants responded to exercise with a decrease in diastolic blood pressure and heart rate, and an improvement in HDL levels, which are some of the different mechanisms through which exercise is known to reduce the risk for T2D and cardiovascular disease [43]. Adipose tissue comprises not only of adipocytes but a mixture of different cell types. To evaluate if the cellular composition of adipose tissue may change during exercise, we looked at the mRNA expression for a number of cell type specific markers before and after the exercise intervention. None of these showed any difference in adipose tissue mRNA expression before vs. after exercise ($q>0.05$; *LEP*, *PNPLA2*, *FAS*, *LIPE* and *PPARG* as markers of adipocytes; *SEBPA/B/D* and *DLK1* as markers of preadipocytes, *PRDM16* and *UCP1* as markers of brown adipocytes; *ITGAX*, *EMR1*, *ITGAM* as markers of macrophages; *TNF* and *IL6* representing cytokines and finally *CCL2* and *CASP7* as markers for inflammation). Although this result suggests that there is no a major change in the cellular composition of the adipose tissue studied before compared with after the exercise intervention, future studies should investigate the methylome in isolated adipocytes. Additionally, in previous studies of DNA methylation in human pancreatic islets, the differences observed in the mixed-cell tissue were also detected in clonal beta cells exposed to hyperglycemia [20,21], suggesting that in at least some tissues, the effects are transferable from the relevant cell type to the tissue of interest for human biology.

The impact of this study is further strengthened by our results showing altered DNA methylation of genes or loci previously associated with obesity and T2D. Although there was no enrichment of differential DNA methylation in those genes compared to the whole dataset, this result may provide a link to the mechanisms for how the loci associated with common diseases exert their functions [18]. 18 obesity and 21 T2D candidate genes had one or more CpG sites which significantly changed in adipose tissue DNA methylation after exercise. 10 CpG sites were found to

Table 2. Changes in adipose tissue DNA methylation in response to a 6 months exercise intervention. Most significant CpG sites ($q<0.005$) with a difference in DNA methylation ≥5%.

	Location in relation to				*DNA Methylation (%)*					
Probe ID	**Chr**	**Nearest Gene**	**Gene region**	**CpG Island**	**Before exercise**	**After exercise**	**Difference**	***p*-value**	***q*-value (<0.005)**	**Cross-reactive probes**
cg04090794	1	*HSP90B3P*	TSS1500	Open sea	31.4±5.1	36.6±4.6	5.2	2.38×10^{-7}	0.004	
cg05091570	1	*NAV1*	Body	CpG Island	30.9±4.1	37.0±3.5	6.1	4.77×10^{-7}	0.004	
cg01828733	1	*NAV1*	TSS200;Body	CpG Island	40.6±4.1	46.2±4.3	5.5	1.19×10^{-6}	0.004	
cg24553673	1	*NR5A2*	Body	S Shore	33.1±4.7	39.9±3.8	6.8	2.38×10^{-7}	0.004	
cg27183818	1		Intergenic	Open sea	66.7±4.7	60.7±4.3	−6.0	7.15×10^{-7}	0.004	
cg26091021	2		Intergenic	N Shelf	38.9±3.6	45.4±3.3	6.5	2.38×10^{-7}	0.004	
cg26297203	2		Intergenic	N Shelf	52.5±3.2	57.6±3.0	5.0	1.19×10^{-6}	0.004	
cg14091208	3	*CCDC48*	Body	CpG Island	41.4±4.6	47.3±4.8	5.9	1.19×10^{-6}	0.004	
cg09217023	3		Intergenic	Open sea	57.2±4.0	62.6±3.2	5.5	2.38×10^{-7}	0.004	
cg09380805	3		Intergenic	N Shelf	29.0±4.0	35.7±3.8	6.6	7.15×10^{-7}	0.004	
cg17103081	4	*GPR125*	Body	N Shelf	63.4±5.2	68.4±4.9	5.1	1.19×10^{-6}	0.004	
cg15133208	4	*SNCA*	5′UTR	N Shore	36.5±4.8	42.4±4.8	6.0	1.67×10^{-6}	0.004	
cg14348967	4		Intergenic	Open sea	31.9±5.2	37.5±4.4	5.6	1.67×10^{-6}	0.004	
cg21817858	5		Intergenic	CpG Island	46.2±5.2	51.8±4.4	5.6	2.38×10^{-7}	0.004	
cg20934416	5		Intergenic	Open sea	76.4±4.9	81.7±3.4	5.3	2.38×10^{-6}	0.005	
cg14246190	6	*EHMT2*	Body	N Shelf	65.1±4.3	70.4±3.4	5.3	2.38×10^{-6}	0.005	
cg20284982	6	*IER3*	TSS1500	S Shore	45.3±5.5	51.2±3.6	5.9	2.38×10^{-6}	0.005	48
cg12586150	6	*SERPINB1*	Body	N Shore	51.9±5.2	58.4±4.7	6.5	2.38×10^{-6}	0.005	
cg09871057	7	*STX1A*	Body	CpG Island	52.3±3.5	57.4±3.2	5.1	1.19×10^{-6}	0.004	
cg18550262	7		Intergenic	Open sea	39.5±3.4	45.0±3.2	5.5	2.38×10^{-7}	0.004	
cg00555695	8	*PVT1*	Body	Open sea	40.3±3.9	45.8±3.4	5.5	2.38×10^{-7}	0.004	48
cg13832372	9	*LHX6*	Body	S Shore	25.8±4.5	31.1±5.4	5.4	2.38×10^{-6}	0.005	
cg02725718	10	*ENKUR*	Body	Open sea	65.6±3.8	70.8±3.1	5.2	2.38×10^{-6}	0.005	47
cg12127706	11	*CTTN*	Body	Open sea	54.3±3.9	59.5±3.7	5.2	1.19×10^{-6}	0.004	
cg02093168	11	*HCCA2*	Body	Open sea	61.2±5.8	67.5±4.2	6.4	1.19×10^{-6}	0.004	47
cg22041190	11	*PKNOX2*	5′UTR	S Shore	36.0±4.5	41.0±4.1	5.0	1.67×10^{-6}	0.004	
cg12439006	11		Intergenic	Open sea	64.5±4.2	69.7±3.2	5.2	2.38×10^{-7}	0.004	
cg19896824	11		Intergenic	Open sea	53.8±5.4	60.6±4.1	6.9	2.38×10^{-7}	0.004	
cg21999471	11		Intergenic	Open sea	41.1±5.3	46.7±3.6	5.6	2.38×10^{-6}	0.005	
cg26828839	12	*ANO2*	Body	Open sea	32.5±5.3	39.7±5.5	7.1	1.19×10^{-6}	0.004	
cg13203394	12	*ITPR2*	Body	Open sea	56.8±4.4	63.3±3.2	6.5	4.77×10^{-7}	0.004	
cg26119796	13	*RB1*	Body	S Shore	57.0±4.8	62.4±4.5	5.4	1.67×10^{-6}	0.004	47
cg00808648	14	*PACS2*	TSS1500	N Shore	44.0±4.1	49.3±4.1	5.3	4.77×10^{-7}	0.004	
cg22396498	15	*CRTC3*	Body	Open sea	59.5±4.5	64.6±5.1	5.1	1.19×10^{-6}	0.004	
cg07299078	16	*KIFC3*	Body;5′UTR	Open sea	49.6±4.3	55.9±4.9	6.4	2.38×10^{-7}	0.004	48
cg05797594	16	*MIR1910;C16orf74*	TSS1500;5′UTR	Open sea	51.5±5.1	57.2±2.9	5.6	2.38×10^{-6}	0.005	47
cg05516390	16	*ZFHX3*	5′UTR	N Shelf	41.8±4.4	49.8±4.4	8.0	1.19×10^{-6}	0.004	
cg06078469	17	*MSI2*	Body	S Shore	43.5±3.6	48.8±4.2	5.4	4.77×10^{-7}	0.004	
cg22386583	17	*RPTOR*	Body	Open sea	51.2±3.8	57.0±3.5	5.8	4.77×10^{-7}	0.004	
cg11225357	17		Intergenic	Open sea	45.1±4.1	50.6±3.9	5.5	1.19×10^{-6}	0.004	
cg20811236	18		Intergenic	N Shore	60.9±5.3	68.2±4.9	7.3	4.77×10^{-7}	0.004	
cg21685776	18		Intergenic	S Shore	51.4±4.4	56.6±4.8	5.2	7.15×10^{-7}	0.004	
cg21520111	19	*TRPM4*	Body	CpG Island	53.4±3.4	59.0±4.0	5.5	4.77×10^{-7}	0.004	
cg21427956	20	*C20orf160*	3′UTR	S Shore	37.5±3.8	43.1±4.3	5.6	1.19×10^{-6}	0.004	
cg08587504	20	*LOC647979*	TSS1500	S Shore	62.7±3.4	68.0±3.0	5.3	2.38×10^{-6}	0.005	
cg10854441	22	*MLC1*	TSS1500	N Shelf	51.3±4.9	57.1±4.3	5.9	1.67×10^{-6}	0.004	

Table 2. Cont.

	Location in relation to				*DNA Methylation (%)*					
Probe ID	**Chr**	**Nearest Gene**	**Gene region**	**CpG Island**	**Before exercise**	**After exercise**	**Difference**	***p*-value**	***q*-value** *(<0.005)*	**Cross-reactive probes**
cg04065832	X	*CDX4*	1stExon	CpG Island	50.8±4.4	56.8±4.6	6.0	2.38×10^{-6}	0.005	
cg19926635	X	*KCND1*	3′UTR	S Shelf	49.8±4.4	55.2±3.9	5.4	1.19×10^{-6}	0.004	

Data are presented as mean ± SD, based on paired non-parametric test and two-tailed *p*-values. Cross-reactive probes: Maximum number of bases (≥47) matched to cross-reactive target as reported by Chen et al. [27].
doi:10.1371/journal.pgen.1003572.t002

have altered DNA methylation in response to exercise within the gene body of *KCNQ1*, a gene encoding a potassium channel and known to be involved in the pathogenesis of T2D, and also subject to parental imprinting [45]. Moreover, exercise associated with changes in DNA methylation of six intragenic CpG sites in *TCF7L2*, the T2D candidate gene harbouring a common variant

Table 3. Changes in adipose tissue DNA methylation in response to a 6 months exercise intervention. Significant CpG sites ($q<0.05$) with the biggest change in DNA methylation (>8%).

	Location in relation to				*DNA Methylation (%)*					
Probe ID	**Chr**	**Nearest Gene**	**Gene**	**CpG Island**	**Before exercise**	**After exercise**	**Difference (>8%)**	***p*-value**	***q*-value**	**Cross-reactive probes**
cg06550177	7		Intergenic	S Shore	29.6±7.2	40.6±7.8	10.9	1.67×10^{-5}	0.008	
cg13906823	1	*TSTD1*	TSS200	CpG Island	39.2±12.5	50.1±15.6	10.9	4.03×10^{-5}	0.011	
cg23397147	17		Intergenic	Open sea	48.1±11.0	58.9±7.5	10.8	4.75×10^{-4}	0.028	
cg24161057	1	*TSTD1*	TSS200	CpG Island	35.9±13.5	46.6±14.6	10.7	2.10×10^{-5}	0.009	
cg26155520	1		Intergenic	Open sea	55.6±7.1	66.0±6.6	10.4	7.87×10^{-6}	0.007	
cg05874882	4		Intergenic	N Shore	34.0±9.1	44.2±6.7	10.1	6.03×10^{-5}	0.013	
cg00257920	1		Intergenic	S Shelf	47.5±9.7	57.5±7.7	10.0	1.53×10^{-4}	0.018	
cg03878654	16	*ZFHX3*	5′UTR	N Shore	56.6±6.7	65.9±6.9	9.3	1.81×10^{-4}	0.019	
cg08360726	19	*PLD3*	5′UTR	CpG Island	29.7±8.0	38.9±11.8	9.2	1.28×10^{-3}	0.043	
cg26682335	17	*ABR*	Body	Open sea	60.6±9.4	69.7±7.0	9.1	2.53×10^{-4}	0.022	
cg01425666	7		Intergenic	CpG Island	33.3±6.8	42.3±5.7	9.0	2.62×10^{-5}	0.010	
cg01750221	12		Intergenic	Open sea	52.3±7.5	61.1±6.4	8.8	8.49×10^{-4}	0.036	
cg05455393	X	*FHL1*	TSS1500	N Shore	52.5±8.4	61.1±7.2	8.6	1.28×10^{-4}	0.017	
cg22828884	3	*FOXP1*	Body	Open sea	62.6±4.4	71.2±4.3	8.6	1.67×10^{-5}	0.008	
cg11837417	19	*CLDND2*	TSS1500	S Shore	65.3±6.4	73.9±5.2	8.6	4.08×10^{-4}	0.027	
cg10323490	2	*THNSL2*	TSS1500	N Shore	64.1±8.1	72.6±6.2	8.5	9.76×10^{-4}	0.038	
cg03934443	10		Intergenic	Open sea	67.4±11.8	75.8±5.6	8.4	9.76×10^{-4}	0.038	
cg01775802	14	*RGS6*	Body	Open sea	63.2±10.1	71.4±10.9	8.2	9.76×10^{-4}	0.038	
cg24606240	1	*NUCKS1*	TSS1500	S Shore	55.4±7.9	63.6±5.9	8.2	7.38×10^{-4}	0.034	
cg23499846	17	*KIAA0664*	5′UTR	S Shore	54.0±5.9	62.0±4.3	8.0	1.03×10^{-5}	0.007	
cg21821308	2	*ASAP2*	Body	CpG Island	42.0±8.5	33.8±5.9	−8.1	3.49×10^{-4}	0.025	
cg19219423	10	*PRKG1*	Body	Open sea	55.4±7.7	47.1±6.8	−8.3	1.81×10^{-4}	0.019	
cg03862437	3	*TMEM44*	Body	N Shore	46.3±7.0	38.0±5.2	−8.3	5.96×10^{-6}	0.006	
cg08368520	7	*FOXK1*	Body	Open sea	52.9±7.8	44.5±8.0	−8.4	9.76×10^{-4}	0.038	
cg01275887	7	*FOXK1*	Body	Open sea	66.3±8.5	57.7±6.6	−8.5	7.38×10^{-4}	0.034	
cg06443678	17		Intergenic	Open sea	51.7±8.2	43.0±6.7	−8.7	2.98×10^{-4}	0.024	
cg02514003	2		Intergenic	Open sea	70.6±6.5	61.7±8.6	−8.9	2.53×10^{-4}	0.022	
cg26504110	19	*LTBP4*	Body	CpG Island	36.9±8.7	27.4±5.1	−9.5	2.98×10^{-4}	0.024	

Data are presented as mean ± SD, based on paired non-parametric test and two-tailed *p*-values. Cross-reactive probes: Maximum number of bases (≥47) matched to cross-reactive target as reported by Chen et al. [27].
doi:10.1371/journal.pgen.1003572.t003

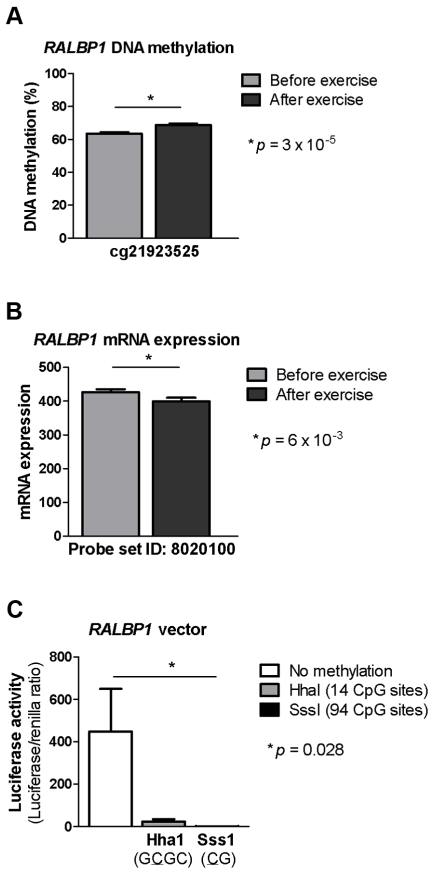


Figure 4. DNA methylation of RALBP1 is associated with a decrease in gene expression. A CpG site in the promoter region of *RALBP1* showed A) increased DNA methylation in response to exercise as well as B) a decrease in mRNA expression. C) *In vitro* DNA methylation of the *RALBP1* promoter decreased gene expression, as measured by luciferase activity. The result represents the mean of three independent experiments, and the values in each experiment are the mean of five replicates (background control subtracted). Data is presented as mean ± SEM.

doi:10.1371/journal.pgen.1003572.g004

with the greatest described effect on the risk of T2D [3]. This is of particular interest considering that *TCF7L2* is subject to alternative splicing [46,47] and the fact that gene exons are more highly methylated than introns, with DNA methylation spikes at splice junctions, suggesting a possible role for differential DNA methylation in transcript splicing [42]. In addition to differential DNA methylation, we also observed an inverse change in adipose tissue mRNA expression for some of these candidate genes, including *TCF7L2*, *HHEX*, *IGF2BP2*, *JAZF1*, *CPEB4* and *SDCCAG8* in response to exercise.

The understanding of the human methylome is incomplete although recently developed methods for genome-wide analysis of DNA methylation already have made, and are likely to continue to make, tremendous advances [48]. High coverage data describing differences in the levels of DNA methylation between certain human tissues or cell types [38], as well as differences observed during development [42], have started to emerge. Regardless, deeper knowledge about the epigenetic architecture and regulation in human adipose tissue has been missing until now. We found that the genetic region with the highest average level of DNA methylation in adipose tissue was the 3′UTR, followed by the gene body and intergenic regions, and those regions also increased the level of DNA methylation in response to exercise. This supports the view that the human methylome can dynamically respond to changes in the environment [14,15]. One explanation for the low average levels of DNA methylation observed in the promoter region (TSS1500/200), 5′UTR and the first exon, may be that these regions often overlap with CpG islands, which are generally known to be unmethylated. Indeed, our results show a very low level of DNA methylation within the CpG islands, and how the level then increases with increasing distances to a CpG island.

It has long been debated if increased DNA methylation precedes gene silencing, or if it is rather a consequence of altered gene activity [40]. The luciferase assay experiments from this study and others [21,23] suggest that DNA methylation may have a causal role, as increased promoter DNA methylation leads to reduced transcriptional activity. Here we further related our findings of altered DNA methylation to mRNA expression, and we identified 197 genes where both DNA methylation and mRNA expression significantly changed in adipose tissue after exercise. Of these, 115 genes (58%) showed an inverse relation, 97% showing an increase in the level of DNA methylation and a decrease in mRNA expression. It should be noted that epigenetic processes are likely to influence more aspects of gene expression, including accessibility of the gene, posttranscriptional RNA processing and stability, splicing and also translation [49]. For example, DNA methylation within the gene body has previously been linked to active gene transcription, suggestively by improving transcription efficiency [42].

Two genes, *HDAC4* and *NCOR2*, with biological relevance in adipose tissue metabolism were selected for functional validation. HDAC4 is a histone deacetylase regulated by phosphorylation, and known to repress GLUT4 transcription in adipocytes [35]. In skeletal muscle, HDAC4 has been found to be exported from the nucleus during exercise, suggesting that removal of the transcriptional repressive function could be a mechanism for exercise adaptation [50]. For *HDAC4*, we observed increased levels of DNA methylation and a simultaneous decrease in mRNA expression in adipose tissue in response to the exercise intervention. Additionally, the functional experiments in cultured adipocytes suggested increased lipogenesis when *Hdac4* expression was reduced. This could be an indicator of reduced repressive activity on GLUT4, leading to an increase in adipocyte glucose uptake and subsequent incorporation of glucose into triglycerides in the process of lipogenesis. *NCOR2* also exhibited increased levels of DNA methylation and a simultaneous decrease in mRNA expression in adipose tissue in response to the exercise intervention, and furthermore we observed increased lipogenesis when *Ncor2* expression was down regulated in the 3T3-L1 cell line. NCOR2 is a nuclear co-repressor, involved in the regulation of genes important for adipogenesis and lipid metabolism, and with the ability to recruit different histone deacetylase enzymes, including HDAC4 [51]. These results may be of clinical importance, since HDAC inhibitors have been suggested in the treatment of obesity and T2D [18,52].

Table 4. Individual CpG sites located within/near candidate genes for obesity [3], with a significant change in DNA methylation in adipose tissue in response to exercise.

Obesity candidate genes			*DNA methylation (%)*						*mRNA expression*				
Probe ID	**Nearest Gene**	**Location**	**Before exercise**	**After exercise**	**Difference**	***p*-value**	***q*-value**	**Cross-reactive probes**	**Before exercise**	**After exercise**	**Difference**	***p*-value**	***q*-value**
cg05501868	*ADAMTS9*	Body	62.3±4.1	66.5±4.5	4.1	6×10^{-4}	0.039		218.5±46.5	219.7±59.8	1.3	>0.05	
cg07233933	*CPEB4*	Body	76.9±2.1	79.2±1.7	2.3	1×10^{-4}	0.013		377.9±63.9	319.3±51.1	−58.7	$<1\times10^{-5}$	0.002
cg09141413	*GPRC5B*	TSS200; CpG Island	15.9±2.4	13.7±1.8	−2.2	6×10^{-4}	0.039		476.3±69.3	448.5±58.4	−27.8	>0.05	
cg22380033	*GRB14*	1stExon; 5′UTR; CpG Island	2.1±0.4	2.6±0.4	0.5	3×10^{-4}	0.029		65.1±11.4	63.0±16.8	−2.1	>0.05	
cg07645296	*ITPR2*	TSS1500; S Shore	56.2±4.5	60.1±4.6	3.9	1×10^{-4}	0.013		421.5±63.1	401.6±91.0	−19.9	>0.05	
cg13203394	*ITPR2*	Body	56.8±4.4	63.3±3.2	6.5	5×10^{-7}	0.001		421.5±63.1	401.6±91.0	−19.9	>0.05	
cg02212836	*LY86*	1stExon	40.0±2.4	42.7±2.5	2.7	9×10^{-5}	0.013		44.2±17.0	57.8±41.0	13.6	>0.05	
cg05021589	*LY86*	TSS200	39.7±2.9	43.5±3.0	3.7	3×10^{-5}	0.013		44.2±17.0	57.8±41.0	13.6	>0.05	
cg09249494	*LY86*	TSS200	31.4±3.9	35.2±3.4	3.9	1×10^{-4}	0.013		44.2±17.0	57.8±41.0	13.6	>0.05	
cg16681597	*LYPLAL1*	Body; CpG Island	9.6±1.6	11.1±2.1	1.6	5×10^{-4}	0.038		146.7±34.7	165.0±28.7	18.3	0.019	0.08
cg01362115	*MAP2K5*	Body	76.3±2.4	78.2±2.1	2.0	7×10^{-4}	0.043		205.2±23.8	195.0±24.6	−10.2	>0.05	
cg02328326	*MAP2K5*	Body	79.6±5.8	84.8±3.8	5.2	6×10^{-4}	0.043		205.2±23.8	195.0±24.6	−10.2	>0.05	
cg20055861	*MAP2K5*	Body	67.4±3.9	71.9±2.9	4.5	1×10^{-4}	0.013		205.2±23.8	195.0±24.6	−10.2	>0.05	
cg27519910	*MSRA*	Body; S Shelf	76.0±2.3	78.6±2.3	2.7	4×10^{-5}	0.013	48	120.4±12.5	124.0±13.5	3.6	>0.05	
cg20147645	*MTIF3*	5′UTR; N Shore	60.0±3.3	64.1±3.6	4.0	3×10^{-4}	0.024		331.3±42.3	331.7±39.8	0.4	>0.05	
cg16420308	*NRXN3*	Body; N Shore	87.3±2.4	89.5±1.7	2.3	7×10^{-4}	0.043		34.4±4.1	34.7±5.4	0.3	>0.05	
cg16592301	*PRKD1*	Body	84.0±2.7	86.3±2.4	2.3	8×10^{-4}	0.048		182.1±20.5	177.4±28.0	−4.7	>0.05	
cg16104450	*SDCCAG8*	Body; N Shore	40.5±3.9	44.2±3.6	3.7	1×10^{-4}	0.013		258.8±22.1	234.9±32.9	−23.9	1×10^{-3}	0.008
cg08222913	*STAB1*	Body; CpG Island	63.2±5.1	67.8±4.2	4.6	2×10^{-5}	0.013		224.1±56.3	214.6±44.2	−9.6	>0.05	
cg26104752	*TBX15*	5′UTR; S Shore	5.8±1.4	6.8±1.5	1.0	7×10^{-5}	0.013		374.1±37.2	368.1±50.9	−6.0	>0.05	
cg19694781	*TMEM160*	Body; N Shore	56.1±6.5	58.3±7.0	2.3	1×10^{-4}	0.013	48	205.0±25.6	231.1±25.7	26.1	1×10^{-3}	0.008
cg05003666	*TUB*	TSS200;Body;CpG Island	21.8±2.5	18.4±2.3	−3.4	3×10^{-4}	0.029		73.6±6.7	72.7±8.0	−0.9	>0.05	
cg01610165	*ZNF608*	Body	10.5±2.3	13.1±2.3	2.6	7×10^{-4}	0.043		171.2±22.9	162.7±25.1	−8.5	>0.05	
cg12817840	*ZNF608*	Body	21.8±2.5	25.8±3.8	3.9	9×10^{-5}	0.013		171.2±22.9	162.7±25.1	−8.5	>0.05	

Data are presented as mean ± SD, based on paired non-parametric test (DNA methylation) or t-test (mRNA expression) and two-tailed *p*-values. Cross-reactive probes: Maximum number of bases (≥47) matched to cross-reactive target as reported by Chen et al. [27].

doi:10.1371/journal.pgen.1003572.t004

Table 5. Individual CpG sites located within/near candidate genes for T2D [3], with a significant change in DNA methylation in adipose tissue in response to exercise.

Type 2 diabetes candidate genes			*DNA methylation (%)*						*mRNA expression*				
Probe ID	**Nearest Gene**	**Location**	**Before exercise**	**After exercise**	**Difference**	***p*-value**	***q*-value**	**Cross-reactive probes**	**Before exercise**	**After exercise**	**Difference**	***p*-value**	***q*-value**
cg05501868	*ADAMTS9*	Body	62.3±4.1	66.5±4.5	4.1	6×10^{-4}	0.025		218.5±46.5	219.7±59.8	1.3	>0.05	
cg21527616	*ADAMTS9*	Body	63.6±4.7	67.0±3.3	3.4	1×10^{-3}	0.044		218.5±46.5	219.7±59.8	1.3	>0.05	
cg14567877	*ADCY5*	Body	80.7±4.3	84.2±3.8	3.5	2×10^{-4}	0.015		257.2±48.3	253.1±44.8	−4.1	>0.05	
cg03720898	*ARAP1*	Body	73.7±3.7	77.1±2.3	3.4	9×10^{-5}	0.013	48	209.4±35.0	202.8±36.4	−6.6	>0.05	
cg06838038	*ARAP1*	Body	42.4±3.9	46.1±3.6	3.7	4×10^{-5}	0.011		209.4±35.0	202.8±36.4	−6.6	>0.05	
cg10495997	*ARAP1*	5′UTR; S Shore	61.8±3.2	64.0±2.9	2.2	2×10^{-4}	0.015		209.4±35.0	202.8±36.4	−6.6	>0.05	
cg15279866	*ARAP1*	5′UTR; Body	56.4±3.5	58.8±3.9	2.4	1×10^{-3}	0.042		209.4±35.0	202.8±36.4	−6.6	>0.05	
cg27058763	*ARAP1*	Body; S Shelf	57.1±4.0	60.7±3.3	3.5	1×10^{-3}	0.041		209.4±35.0	202.8±36.4	−6.6	>0.05	
cg01865786	*BCL11A*	Body	64.7±4.3	67.7±2.8	3.0	8×10^{-4}	0.034		19.4±2.1	21.3±2.7	1.8	0.009	0.04
cg03390300	*CDKAL1*	Body	85.2±2.3	87.5±1.8	2.4	2×10^{-5}	0.011		263.1±24.7	268.3±25.1	5.2	>0.05	
cg07562918	*CDKN2A*	1stExon; CpG Island	16.7±2.1	18.4±2.2	1.6	1×10^{-3}	0.042		35.9±5.3	40.5±7.0	4.6	0.023	0.09
cg20836993	*DGKB*	Body	68.3±1.8	70.0±1.7	1.7	6×10^{-4}	0.025	48	16.8±3.5	17.8±3.9	0.9	>0.05	
cg01602287	*DUSP8*	Body; CpG Island	75.4±4.5	79.5±3.4	4.2	1×10^{-3}	0.042		97.7±13.9	94.9±13.5	−2.9	>0.05	
cg26902557	*DUSP8*	Body	49.1±3.9	52.1±4.0	3.0	6×10^{-4}	0.028		97.7±13.9	94.9±13.5	−2.9	>0.05	
cg26580413	*FTO*	Body	61.0±4.4	64.3±3.8	3.3	1×10^{-3}	0.044		785.0±80.7	794.4±64.7	9.4	>0.05	
cg20180364	*HHEX*	TSS1500; N Shore	46.8±4.1	50.4±3.5	3.6	2×10^{-4}	0.015		172.7±24.7	144.9±30.3	−27.8	5×10^{-5}	<0.001
cg16965605	*HMGA2*	Body	70.2±5.6	75.1±4.3	4.9	1×10^{-3}	0.037		32.2±3.2	34.9±3.6	2.7	0.004	0.025
cg17182048	*HMGA2*	Body	81.2±3.9	84.5±4.8	3.4	1×10^{-3}	0.044		32.2±3.2	34.9±3.6	2.7	0.004	0.025
cg17518348	*HMGA2*	Body	78.8±3.7	83.3±3.5	4.5	2×10^{-4}	0.015		32.2±3.2	34.9±3.6	2.7	0.004	0.025
cg06150454	*IGF2BP2*	Body	54.6±4.2	58.4±2.8	3.8	3×10^{-4}	0.021	48	105.6±16.5	88.4±14.9	−17.1	$<1\times10^{-5}$	<0.001
cg13918631	*IGF2BP2*	Body	66.8±4.7	70.7±3.8	3.9	2×10^{-4}	0.015		105.6±16.5	88.4±14.9	−17.1	$<1\times10^{-5}$	<0.001
cg02963803	*JAZF1*	Body	59.6±2.6	62.1±2.5	2.5	9×10^{-5}	0.013	48	238.2±26.1	218.5±28.0	−19.7	0.01	0.047
cg01689159	*KCNQ1*	Body; CpG Island	80.5±2.6	83.3±1.7	2.7	3×10^{-4}	0.017		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg03660952	*KCNQ1*	Body	51.8±3.5	55.0±2.6	3.2	1×10^{-5}	0.011		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg04894537	*KCNQ1*	Body	40.5±3.5	44.6±4.3	4.2	3×10^{-4}	0.021		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg06838584	*KCNQ1*	Body	46.8±3.7	44.0±3.4	−2.7	2×10^{-4}	0.015		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg08160246	*KCNQ1*	Body	60.3±3.3	63.5±3.3	3.2	5×10^{-4}	0.025		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg13577072	*KCNQ1*	Body	67.4±3.1	71.8±3.5	4.5	6×10^{-5}	0.011		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg15910264	*KCNQ1*	Body	81.4±2.8	84.3±1.9	2.9	3×10^{-5}	0.011		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg19672982	*KCNQ1*	Body	70.4±3.0	73.3±2.8	2.9	1×10^{-4}	0.014		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg24725201	*KCNQ1*	Body	91.9±1.6	93.4±1.4	1.5	7×10^{-4}	0.031		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg25786675	*KCNQ1*	Body	66.3±3.7	62.4±3.6	−3.9	6×10^{-5}	0.011		67.0±7.0	66.1±7.4	−0.9	>0.05	
cg04775232	*PRC1*	Body	82.1±2.4	84.0±2.2	1.9	2×10^{-3}	0.048		64.3±12.0	59.8±15.8	−4.5	>0.05	

Table 5. Cont.

Type 2 diabetes candidate genes			*DNA methylation (%)*						*mRNA expression*				
Probe ID	**Nearest Gene**	**Location**	**Before exercise**	**After exercise**	**Difference**	***p*-value**	***q*-value**	**Cross-reactive probes**	**Before exercise**	**After exercise**	**Difference**	***p*-value**	***q*-value**
cg01902845	*PROX1*	Body	73.5±5.0	77.7±3.4	4.2	6×10^{-4}	0.025		21.2±6.2	21.2±7.0	0	>0.05	
cg14545834	*PTPRD*	Body; CpG Island	68.0±2.8	71.2±2.2	3.2	2×10^{-4}	0.015	49	80.3±17.8	81.8±14.5	1.4	>0.05	
cg00831931	*TCF7L2*	Body	82.4±2.7	84.8±2.5	2.4	2×10^{-4}	0.015		529.9±58.9	474.0±74.7	−55.8	0.001	0.008
cg05923857	*TCF7L2*	Body	72.6±5.2	76.4±3.8	3.8	8×10^{-4}	0.034		529.9±58.9	474.0±74.7	−55.8	0.001	0.008
cg06403317	*TCF7L2*	Body	92.1±2.6	94.2±1.7	2.1	1×10^{-3}	0.037		529.9±58.9	474.0±74.7	−55.8	0.001	0.008
cg09022607	*TCF7L2*	Body; S Shore	25.5±4.5	21.2±3.1	−4.3	6×10^{-4}	0.025		529.9±58.9	474.0±74.7	−55.8	0.001	0.008
cg19226647	*TCF7L2*	1stExon; N Shore	4.4±1.0	5.5±1.3	1.1	4×10^{-5}	0.011		529.9±58.9	474.0±74.7	−55.8	0.001	0.008
cg23951816	*TCF7L2*	Body	63.8±3.8	68.4±3.5	4.6	5×10^{-4}	0.025		529.9±58.9	474.0±74.7	−55.8	0.001	0.008
cg01649611	*THADA*	Body	38.5±5.3	42.5±4.6	4.0	2×10^{-4}	0.015		285.2±29.9	291.7±31.4	6.5	>0.05	
cg12277798	*THADA*	Body; S Shelf	77.2±4.4	81.6±3.3	4.5	5×10^{-4}	0.025		285.2±29.9	291.7±31.4	6.5	>0.05	
cg16417416	*WFS1*	Body	63.9±3.5	66.9±2.9	2.9	1×10^{-3}	0.044		132.2±21.6	123.6±14.5	−8.6	0.036	0.13
cg22051204	*ZBED3*	5′UTR; S Shore	51.1±3.5	53.5±3.2	2.4	1×10^{-3}	0.042		187.9±22.9	189.1±17.4	1.3	>0.05	

Data are presented as mean ± SD, based on paired non-parametric test (DNA methylation) or t-test (mRNA expression) and two-tailed *p*-values. Cross-reactive probes: Maximum number of bases (≥47) matched to cross-reactive target as reported by Chen et al. [27].

doi:10.1371/journal.pgen.1003572.t005

In summary, this study provides a detailed map of the human methylome in adipose tissue, which can be used as a reference for further studies. We have also found evidence for an association between differential DNA methylation and mRNA expression in response to exercise, as well as a connection to genes known to be involved in the pathogenesis of obesity and T2D. Finally, functional validation in adipocytes links DNA methylation via gene expression to altered metabolism, supporting the role of histone deacetylase enzymes as a potential candidate in clinical interventions.

Materials and Methods

Ethics statement

Written informed consent was obtained from all participants and the research protocol was approved by the local human research ethics committee.

Study participants

This study included a total of 31 men from Malmö, Sweden, recruited for a six months exercise intervention study, as previously described [23,53]. Fifteen of the individuals had a first-degree family history of T2D (FH^+), whereas sixteen individuals had no family history of diabetes (FH^-). They were all sedentary, but healthy, with a mean age of 37.4 years and a mean BMI of 27.8 kg/m^2 at inclusion. All subjects underwent a physical examination, an oral glucose tolerance test and a submaximal exercise stress test. Bioimpedance was determined to estimate fat mass with a BIA 101 Body Impedance Analyzer (Akern Srl, Pontassieve, Italy). To directly assess the maximal oxygen uptake (VO_{2max}), an ergometer bicycle (Ergomedic 828E, Monark, Sweden) was used together with heart rate monitoration (Polar T61, POLAR, Finland) [53]. FH^+ and FH^- men were group-wise matched for age, BMI and physical fitness (VO_{2max}) at baseline. Subcutaneous biopsies of adipose tissue from the right thigh were obtained during the fasting state under local anaesthesia (1% Lidocaine) using a 6 mm Bergström needle (Stille AB, Sweden) from all participants before and from 23 participants after the six months exercise intervention (>48 hours after the last exercise session). The weekly group training program included one session of 1 hour spinning and two sessions of 1 hour aerobics and was led by a certified instructor. The participation level was on average 42.8±4.5 sessions, which equals to 1.8 sessions/week of this endurance exercise intervention. The study participants were requested to not change their diet and daily activity level during the intervention.

Genome-wide DNA methylation analysis

DNA methylation was analyzed in DNA extracted from adipose tissue, using the Infinium HumanMethylation450 BeadChip assay (Illumina, San Diego, CA, USA). This array contains 485,577 probes, which cover 21,231 (99%) RefSeq genes [25,54]. Genomic DNA (500 ng) from adipose tissue was bisulfite treated using the EZ DNA methylation kit (Zymo Research, Orange, CA, USA). Analysis of DNA methylation with the Infinium assay was carried out on the total amount of bisulfite-converted DNA, with all other procedures following the standard Infinium HD Assay Methylation Protocol Guide (Part #15019519, Illumina). The BeadChips' images were captured using the Illumina iScan. The raw methylation score for each probe represented as methylation β-values was calculated using GenomeStudio Methylation module software (β = intensity of the Methylated allele (M)/intensity of the Unmethylated allele (U)+intensity of the Methylated allele (M)+100). All included samples showed a high quality bisulfite

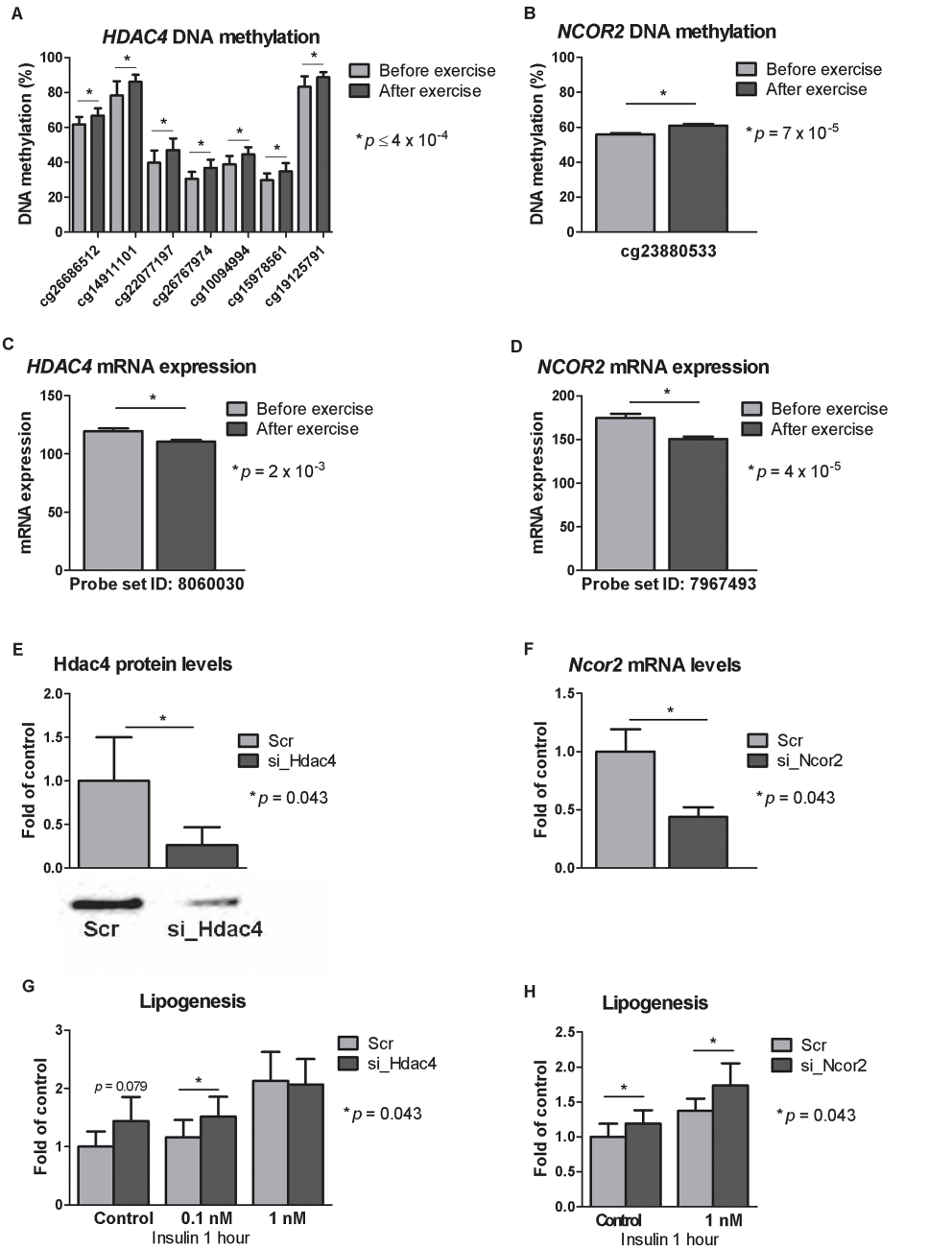

A
HDAC4 DNA methylation
DNA methylation (%)
Before exercise
After exercise
*p ≤ 4 x 10-4
cg26686512
cg14911101
cg22077197
cg26767974
cg10094994
cg15978561
cg19125791
B
NCOR2 DNA methylation
DNA methylation (%)
Before exercise
After exercise
*p = 7 x 10-5
cg23880533
C
HDAC4 mRNA expression
mRNA expression
Before exercise
After exercise
*p = 2 x 10-3
Probe set ID: 8060030
D
NCOR2 mRNA expression
mRNA expression
Before exercise
After exercise
*p = 4 x 10-5
Probe set ID: 7967493
E
Hdac4 protein levels
Fold of control
Scr
si_Hdac4
*p = 0.043
Scr
si_Hdac4
F
Ncor2 mRNA levels
Fold of control
Scr
si_Ncor2
*p = 0.043
G
Lipogenesis
Fold of control
p = 0.079
Scr
si_Hdac4
*p = 0.043
Control
0.1 nM
1 nM
Insulin 1 hour
H
Lipogenesis
Fold of control
Scr
si_Ncor2
*p = 0.043
Control
1 nM
Insulin 1 hour

Figure 5. Silencing of Hdac4 and Ncor2 in 3T3-L1 adipocytes results in increased lipogenesis. CpG sites in the promoter region of A) *HDAC4* and B) *NCOR2* showed increased DNA methylation in response to exercise as well as decreased mRNA expression (C–D). Knock-downs were verified either by E) Western blot analysis (for Hdac4) or F) by qRT-PCR (for *Ncor2*). Lipogenesis increased in 3T3-L1 adipocytes where G) *Hdac4* ($n = 5$) or H) *Ncor2* ($n = 5$) had been silenced. Data is presented as mean ± SEM.
doi:10.1371/journal.pgen.1003572.g005

conversion efficiency (intensity signal >4000) [55], and also passed all GenomeStudio quality control steps based on built in control probes for staining, hybridization, extension and specificity. Individual probes were then filtered based on Illumina detection *p*-value and all CpG sites with a mean $p<0.01$ were considered detected and used for subsequent analysis. In total we obtained DNA methylation data for 476,753 CpG sites from adipose tissue of 31 men before and 23 men after the exercise intervention. Before further analysis, the DNA methylation data was exported from GenomeStudio and subsequently analyzed using Bioconductor [56] and the lumi package [57]. β-values were converted to M-values ($M = \log^2(\beta/(1-\beta))$), a more statistically valid method for conducting differential methylation analysis [58]. Next, data was background corrected by subtracting the median M-value of the 600 built in negative controls and was further normalized using quantile normalization. Correction for batch effects within the methylation array data was performed using COMBAT [59]. For the calculations of global DNA methylation, quantile normalization was omitted and probes reported to be cross-reactive (≥49 bases) or directly affected by a SNP (MAF>5%) were removed [27]. Due to different performance of Infinium I and Infinium II assays [25], the results based on average DNA methylation are calculated and presented separately for each probe type. To control for technical variability within the experiment, one adipose tissue sample was included and run on four different occasions (Figure S1a). As the β-value is easier to interpret biologically, M-values were reconverted to β-values when describing the results and creating the figures.

mRNA expression analysis

RNA extracted from the subcutaneous adipose tissue biopsies was used for a microarray analysis, performed using the GeneChip Human Gene 1.0 ST whole transcript based array (Affymetrix, Santa Clara, CA, USA), following the Affymetrix standard protocol. Basic Affymetrix chip and experimental quality analyses were performed using the Expression Console Software, and the robust multi-array average method (RMA) was used for background correction, data normalization and probe summarization [60].

Luciferase assay

The human promoter fragment containing 1500 bp of DNA upstream of the transcription start site for *RALBP1* (Chr18:9474030–9475529, GRCh37/hg19) was inserted into a CpG-free luciferase reporter vector (pCpGL-basic) as previously described [21]. The construct was methylated using two different DNA methyltransferases; SssI which methylates all cytosine residues within the double-stranded dinucleotide recognition sequence CG, and HhaI which methylates only the internal cytosine residue in the GCGC sequence (New England Biolabs, Frankfurt, Germany). INS-1 cells were co-transfected with 100 ng of the pCpGL-vector without (control) or with any of the three *RALBP1* inserts (no DNA methyltransferase, SssI, HhaI) together with 2 ng of pRL renilla luciferase control reporter vector as a control for transfection efficiency (Promega, Madison, WI, USA). Firefly luciferase activity, as a value of expression, was measured for each construct and normalized against renilla luciferase activity using the TD-20/20 luminometer (Turner Designs, Sunnyvale, CA, USA). The results represent the mean of three independent experiments, and the values in each experiment are the mean of five replicates. Cells transfected with an empty pCpGL-vector were used as background control in each experiment.

siRNA transfection of 3T3-L1 adipocytes and lipogenesis assay

For detailed description of siRNA and lipogenesis experiments see Methods S1. Briefly, 3T3-L1 fibroblasts were cultured at sub-confluence in DMEM containing 10% (v/v) FCS, 100 U/ml penicillin and 100 µg/ml streptomycin at 37°C and 95% air/5% CO_2. Two-day post-confluent cells were incubated for 72 h in DMEM supplemented with 0.5 mM IBMX, 10 µg/ml insulin and 1 µM dexamethasone, after which the cells were cultured in normal growth medium. Seven days post-differentiation, cells were transfected by electroporation with 2 nmol of each siRNA sequence/gene (Table S5). 0.2 nmol scrambled siRNA of each low GC-, medium GC- and high GC-complex were mixed as control. The cells were replated after transfection and incubated for 72 hours (siRNA against *Hdac4*) or 24 hours (siRNA against *Ncor2*).

Cells harvested for western blot analysis were solubilized and homogenized, and 20 µg protein was subjected to SDS-PAGE (4–12% gradient) and subsequent transferred to nitrocellulose membranes. The primary antibody (rabbit polyclonal anti-hdac4; ab12172, Abcam, Cambridge, UK) was diluted in 5 ml 5% BSA/TBST and incubated overnight in 4°C. The secondary antibody (goat anti-rabbit IgG conjugated to horseradish peroxidase; ALI4404, BioSource, Life Technologies Ltd, Paisley, UK) was diluted 1:20,000 in 5% milk/TBST. Protein was detected using Super Signal and ChemiDoc (BioRad, Hercules, CA, USA).

Quantitative PCR (Q-PCR) analyses were performed in triplicate on an ABI7900 using Assays on demand with TaqMan technology (Mm00448796_m1, Applied Biosystems, Carlsbad, CA, USA). The mRNA expression was normalized to the expression of the endogenous control gene *Hprt* (Mm01545399_m1, Applied Biosystems).

To measure lipogenesis, 10 µl tritium labelled ($[^3H]$) glucose (Perkin Elmer, Waltham, MA, USA) was added followed by insulin of different concentrations; 0, 0.1, and 1 nM for *Hdac4* siRNA and 0 and 1 nM for *Ncor2* siRNA experiments, respectively. All concentrations were tested in duplicates. After 1 hour, incorporation of $[^3H]$ glucose into cellular lipids was measured by scintillation counting. Lipogenesis is expressed as fold of basal lipogenesis.

DNA methylation analysis using PyroSequencing

PyroSequencing (PyroMark Q96ID, Qiagen, Hilden, Germany) was used to technically validate data from the genome-wide DNA methylation analysis. PCR and sequencing primers were either designed using PyroMark Assay Design 2.0 or ordered as pre-designed methylation assays (Qiagen, Table S4), and all procedures were performed according to recommended protocols. Briefly, 100 ng genomic DNA from adipose tissue of 23 individuals both before and after the exercise intervention was bisulfite converted using Qiagen's EpiTect kit. With one primer biotinylated at its 5′ end, bisulfite-converted DNA was amplified by PCR using the PyroMark PCR Master Mix kit (Qiagen).

Biotinylated PCR products were immobilized onto streptavidin-coated beads (GE Healthcare, Uppsala, Sweden) and DNA strands were separated using denaturation buffer. After washing and neutralizing using PyroMark Q96 Vacuum Workstation, the sequencing primer was annealed to the immobilized strand. PyroSequencing was performed with the PyroMark Gold Q96 reagents and data were analyzed using the PyroMark Q96 (version 2.5.8) software (Qiagen).

Statistical analysis

Clinical data is presented as mean ± SD, and comparisons based on a t-test and two-tailed *p*-values. Genome-wide DNA methylation data from the Infinium HumanMethylation450 BeadChip before vs. after the six month exercise intervention was analyzed using a paired non-parametric test, whereas a paired t-test was used to compare the mRNA expression. DNA methylation and mRNA expression data are expressed as mean ± SD. To account for multiple testing and reduce the number of false positives, we applied *q*-values to measure the false discovery rate (FDR) on our genome-wide analyses of DNA methylation and mRNA expression [24]. Luciferase activity was analyzed using the Friedman test (paired, non-parametric test on dependent samples) and presented as mean ± SEM. Data from 3T3-L1 adipocyte experiments showing protein, mRNA and lipogenesis levels are presented as mean ± SEM, and the results are based on Wilcoxon signed-rank test.

Supporting Information

Figure S1 Technical validation. A) Technical replicate of one adipose tissue DNA sample included in the study, analyzed using the Infinium HumanMethylation450 BeadChip on four different occasions. B–C) Data obtained from all adipose tissue samples for four CpG sites, from both the Infinium HumanMethylation450 BeadChip (x axis) and using Pyrosequencing (y axis).
(TIF)

Methods S1 Detailed descriptions of small interfering RNA transfection, mRNA expression analysis, lipogenesis assay and statistical analysis.
(DOC)

Table S1 Baseline clinical characteristics of individuals with (FH^+) or without (FH^-) a family history of type 2 diabetes.
(DOC)

Table S2 Average DNA methylation for regions in relation to nearest gene or CpG islands, separately for Infinium I and II assays, respectively.
(DOC)

Table S3 CpG sites with a change in DNA methylation ($q<0.05$ and difference in $\beta \geq 5\%$) concurrent with an inverse change in mRNA expression ($q<0.05$) of the nearest gene, in response to the exercise intervention study.
(DOC)

Table S4 Assay design for technical validation of DNA methylation data using PyroSequencing.
(DOC)

Table S5 siRNA assays.
(DOC)

Acknowledgments

Ylva Wessman is acknowledged for skilled technical assistance, Targ Elgzyri for collection of clinical material and Peter Almgren for advice on the statistical calculations. We acknowledge SCIBLU (Swegene Center for Integrative Biology at Lund University) Genomics Facility for help with DNA methylation and mRNA expression analyses.

Author Contributions

Conceived and designed the experiments: TR PV KFE HAJ LG CL. Performed the experiments: TR CD TD EN ÅT MDN. Analyzed the data: TR PV CD TD EH AHO CL. Contributed reagents/materials/analysis tools: KFE HAJ LG. Wrote the paper: TR PV CD TD EH AHO EN MDN HAJ LG CL.

References

1. Ng SW, Popkin BM (2012) Time use and physical activity: a shift away from movement across the globe. Obes Rev 13: 659–80.
2. Ronti T, Lupattelli G, Mannarino E (2006) The endocrine function of adipose tissue: an update. Clin Endocrinol (Oxf) 64: 355–365.
3. McCarthy MI (2010) Genomics, type 2 diabetes, and obesity. N Engl J Med 363: 2339–2350.
4. Almgren P, Lehtovirta M, Isomaa B, Sarelin L, Taskinen MR, et al. (2011) Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. Diabetologia 54: 2811–2819.
5. Groop L, Forsblom C, Lehtovirta M, Tuomi T, Karanko S, et al. (1996) Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. Diabetes 45: 1585–1593.
6. Isomaa B, Forsen B, Lahti K, Holmstrom N, Waden J, et al. (2010) A family history of diabetes is associated with reduced physical fitness in the Prevalence, Prediction and Prevention of Diabetes (PPP)-Botnia study. Diabetologia 53: 1709–1713.
7. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, et al. (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 346: 393–403.
8. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, et al. (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 344: 1343–1350.
9. Gluckman PD, Hanson MA, Buklijas T, Low FM, Beedle AS (2009) Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. Nat Rev Endocrinol 5: 401–408.
10. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, et al. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102: 10604–10609.
11. Ling C, Poulsen P, Simonsson S, Ronn T, Holmkvist J, et al. (2007) Genetic and epigenetic factors are associated with expression of respiratory chain component NDUFB6 in human skeletal muscle. J Clin Invest 117: 3427–3435.
12. Ronn T, Poulsen P, Hansson O, Holmkvist J, Almgren P, et al. (2008) Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle. Diabetologia 51: 1159–1168.
13. Sandovici I, Smith NH, Nitert MD, Ackers-Johnson M, Uribe-Lewis S, et al. (2011) Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets. Proc Natl Acad Sci U S A 108: 5449–5454.
14. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6–21.
15. Feinberg AP, Irizarry RA (2010) Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci U S A 107 Suppl 1: 1757–1764.
16. Barres R, Osler ME, Yan J, Rune A, Fritz T, et al. (2009) Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. Cell Metab 10: 189–198.
17. Ling C, Del Guerra S, Lupi R, Ronn T, Granhall C, et al. (2008) Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia 51: 615–622.
18. Ling C, Groop L (2009) Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes 58: 2718–2725.
19. Volkmar M, Dedeurwaerder S, Cunha DA, Ndlovu MN, Defrance M, et al. (2012) DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. EMBO J 31: 1405–1426.
20. Yang BT, Dayeh TA, Kirkpatrick CL, Taneera J, Kumar R, et al. (2011) Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. Diabetologia 54: 360–367.
21. Yang BT, Dayeh TA, Volkov PA, Kirkpatrick CL, Malmgren S, et al. (2012) Increased DNA Methylation and Decreased Expression of PDX-1 in Pancreatic Islets from Patients with Type 2 Diabetes. Mol Endocrinol 26: 1203–12.

22. Bouchard L, Rabasa-Lhoret R, Faraj M, Lavoie ME, Mill J, et al. (2010) Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction. Am J Clin Nutr 91: 309–320.
23. Nitert MD, Dayeh T, Volkov P, Elgzyri T, Hall E, et al. (2012) Impact of an Exercise Intervention on DNA Methylation in Skeletal Muscle From First-Degree Relatives of Patients With Type 2 Diabetes. Diabetes 61: 3322–32.
24. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100: 9440–9445.
25. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, et al. (2011) High density DNA methylation array with single CpG site resolution. Genomics 98: 288–295.
26. Maksimovic J, Gordon L, Oshlack A (2012) SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol 13: R44.
27. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, et al. (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8: 203–9.
28. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, et al. (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet 8: e1002629.
29. Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, et al. (2011) Epigenetic predictor of age. PLoS One 6: e14821.
30. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, et al. (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res 20: 434–439.
31. Singhal J, Nagaprashantha L, Vatsyayan R, Awasthi S, Singhal SS (2011) RLIP76, a glutathione-conjugate transporter, plays a major role in the pathogenesis of metabolic syndrome. PLoS One 6: e24688.
32. Chen XW, Leto D, Chiang SH, Wang Q, Saltiel AR (2007) Activation of RalA is required for insulin-stimulated Glut4 trafficking to the plasma membrane via the exocyst and the motor protein Myo1c. Dev Cell 13: 391–404.
33. Fang S, Suh JM, Atkins AR, Hong SH, Leblanc M, et al. (2011) Corepressor SMRT promotes oxidative phosphorylation in adipose tissue and protects against diet-induced obesity and insulin resistance. Proc Natl Acad Sci U S A 108: 3412–3417.
34. Sutanto MM, Ferguson KK, Sakuma H, Ye H, Brady MJ, et al. (2010) The silencing mediator of retinoid and thyroid hormone receptors (SMRT) regulates adipose tissue accumulation and adipocyte insulin sensitivity in vivo. J Biol Chem 285: 18485–18495.
35. Weems JC, Griesel BA, Olson AL (2012) Class II histone deacetylases downregulate GLUT4 transcription in response to increased cAMP signaling in cultured adipocytes and fasting mice. Diabetes 61: 1404–1414.
36. Barres R, Yan J, Egan B, Treebak JT, Rasmussen M, et al. (2012) Acute exercise remodels promoter methylation in human skeletal muscle. Cell Metab 15: 405–411.
37. Zhang FF, Cardarelli R, Carroll J, Zhang S, Fulda KG, et al. (2011) Physical activity and global genomic DNA methylation in a cancer-free population. Epigenetics 6: 293–299.
38. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38: 1378–1385.
39. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 41: 178–186.
40. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13: 484–492.
41. Dayeh TA, Olsson AH, Volkov P, Almgren P, Ronn T, et al. (2013) Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia 56: 1036–46.
42. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, et al. (2010) Dynamic changes in the human methylome during differentiation. Genome Res 20: 320–331.
43. Slentz CA, Houmard JA, Kraus WE (2009) Exercise, abdominal obesity, skeletal muscle, and metabolic risk: evidence for a dose response. Obesity (Silver Spring) 17 Suppl 3: S27–33.
44. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, et al. (2010) Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet 42: 949–960.
45. Travers ME, Mackay DJ, Nitert MD, Morris AP, Lindgren CM, et al. (2012) Insights Into the Molecular Mechanism for Type 2 Diabetes Susceptibility at the KCNQ1 Locus From Temporal Changes in Imprinting Status in Human Islets. Diabetes 62: 987–92.
46. Kaminska D, Kuulasmaa T, Venesmaa S, Kakela P, Vaittinen M, et al. (2012) Adipose Tissue TCF7L2 Splicing Is Regulated by Weight Loss and Associates With Glucose and Fatty Acid Metabolism. Diabetes 61: 2807–2813.
47. Osmark P, Hansson O, Jonsson A, Ronn T, Groop L, et al. (2009) Unique splicing pattern of the TCF7L2 gene in human pancreatic islets. Diabetologia 52: 850–854.
48. Emes RD, Farrell WE (2012) Make way for the 'next generation': application and prospects for genome-wide, epigenome-specific technologies in endocrine research. J Mol Endocrinol 49: R19–27.
49. Gibney ER, Nolan CM (2010) Epigenetics and gene expression. Heredity (Edinb) 105: 4–13.
50. McGee SL, Fairlie E, Garnham AP, Hargreaves M (2009) Exercise-induced histone modifications in human skeletal muscle. J Physiol 587: 5951–5958.
51. Watson PJ, Fairall L, Schwabe JW (2012) Nuclear hormone receptor co-repressors: structure and function. Mol Cell Endocrinol 348: 440–449.
52. Galmozzi A, Mitro N, Ferrari A, Gers E, Gilardi F, et al. (2012) Inhibition of Class I Histone Deacetylases Unveils a Mitochondrial Signature and Enhances Oxidative Metabolism in Skeletal Muscle and Adipose Tissue. Diabetes 62: 732–42.
53. Elgzyri T, Parikh H, Zhou Y, Nitert MD, Ronn T, et al. (2012) First-Degree Relatives of Type 2 Diabetic Patients Have Reduced Expression of Genes Involved in Fatty Acid Metabolism in Skeletal Muscle. J Clin Endocrinol Metab 97: E1332–7.
54. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, et al. (2011) Evaluation of the Infinium Methylation 450K technology. Epigenomics 3: 771–784.
55. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, et al. (2009) An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One 4: e8274.
56. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.
57. Du P, Kibbe WA, Lin SM (2008) lumi: a pipeline for processing Illumina microarray. Bioinformatics 24: 1547–1548.
58. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, et al. (2010) Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11: 587.
59. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8: 118–127.
60. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249–264.

Study II

OXFORD

Human Molecular Genetics, 2015, Vol. 24, No. 13 3792–3813

doi: 10.1093/hmg/ddv124
Advance Access Publication Date: 10 April 2015
Original Article

ORIGINAL ARTICLE

Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood

Tina Rönn[1,†], Petr Volkov[1,†], Linn Gillberg[3,4,†], Milana Kokosar[5], Alexander Perfilyev[1], Anna Louisa Jacobsen[3], Sine W. Jørgensen[3,6], Charlotte Brøns[3], Per-Anders Jansson[7], Karl-Fredrik Eriksson[8], Oluf Pedersen[9], Torben Hansen[9], Leif Groop[2], Elisabet Stener-Victorin[5,10], Allan Vaag[3,4], Emma Nilsson[1,3,‡] and Charlotte Ling[1,*,‡]

[1]Department of Clinical Sciences, Epigenetics and Diabetes and [2]Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, CRC, 205 02 Malmö, Sweden, [3]Department of Endocrinology, Rigshospitalet, Tagensvej 20, DK-2200 Copenhagen, Denmark, [4]Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen, Denmark, [5]Department of Physiology/Endocrinology, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Medicinaregatan 11, Box 434, 405 30 Gothenburg, Sweden, [6]Steno Diabetes Center, Niels Steensensvej 2, DK-2820 Gentofte, Denmark, [7]Wallenberg Laboratory, Sahlgrenska University Hospital, Gothenburg, Sweden, [8]Department of Clinical Sciences, Vascular Diseases, Lund University, 205 02 Malmö, Sweden, [9]The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Universitetsparken 1, 2100 Copenhagen, Denmark and [10]Department of Physiology and Pharmacology, Karolinska Institutet, 171 77 Stockholm, Sweden

*To whom correspondence should be addressed at: Department of Clinical Sciences, Epigenetic and Diabetes Unit, Lund University Diabetes Centre, Jan Waldenströms gata 35, Clinical Research Centre Building 91, Level 12, 205 02 Malmö, Sweden. Tel: +46 40391213; Fax: +46 40391222; Email: charlotte.ling@med.lu.se

Abstract

Increased age, BMI and HbA1c levels are risk factors for several non-communicable diseases. However, the impact of these factors on the genome-wide DNA methylation pattern in human adipose tissue remains unknown. We analyzed the DNA methylation of ∼480 000 sites in human adipose tissue from 96 males and 94 females and related methylation to age, BMI and HbA1c. We also compared epigenetic signatures in adipose tissue and blood. Age was significantly associated with both altered DNA methylation and expression of 1050 genes (e.g. *FHL2*, *NOX4* and *PLG*). Interestingly, many reported epigenetic biomarkers of

† These authors contributed equally to this study.
‡ These authors contributed equally to this study.
Received: January 29, 2015. **Revised and Accepted:** April 7, 2015

aging in blood, including *ELOVL2*, *FHL2*, *KLF14* and *GLRA1*, also showed significant correlations between adipose tissue DNA methylation and age in our study. The most significant association between age and adipose tissue DNA methylation was found upstream of *ELOVL2*. We identified 2825 genes (e.g. *FTO*, *ITIH5*, *CCL18*, *MTCH2*, *IRS1* and *SPP1*) where both DNA methylation and expression correlated with BMI. Methylation at previously reported *HIF3A* sites correlated significantly with BMI in females only. HbA1c (range 28–46 mmol/mol) correlated significantly with the methylation of 711 sites, annotated to, for example, *RAB37*, *TICAM1* and *HLA-DPB1*. Pathway analyses demonstrated that methylation levels associated with age and BMI are overrepresented among genes involved in cancer, type 2 diabetes and cardiovascular disease. Our results highlight the impact of age, BMI and HbA1c on epigenetic variation of candidate genes for obesity, type 2 diabetes and cancer in human adipose tissue. Importantly, we demonstrate that epigenetic biomarkers in blood can mirror age-related epigenetic signatures in target tissues for metabolic diseases such as adipose tissue.

Introduction

Epigenetic factors, including DNA methylation, histone modifications and various RNA-mediated processes, are involved in tissue-specific gene regulation and have been suggested as mechanisms for interaction between environmental factors and the genome (1). As epigenetic variation affects genome function, it may also contribute to common human diseases (2). Indeed, a number of factors support the involvement of epigenetic components in common complex diseases, e.g. monozygotic twins do not show 100% concordance for common diseases and indeed display epigenetic differences (3–6), the incidence of several complex diseases is rising in the general population (7), and there is an association between *in utero* environment or early development and diseases in adult life (8–10). DNA methylation is an easily accessible epigenetic mark for laboratory investigations and thereby suited for epigenome-wide association studies (EWAS) and may be used as an epigenetic biomarker (2). However, the fact that epigenetic alterations may be either causal or arise as a consequence of disease needs to be accounted for. It is hence important to study the impact of non-genetic risk factors for disease, e.g. age, BMI and HbA1c (a measure of long-term glycemia) (11–15), on epigenetic modifications prior to disease development. These three non-genetic risk factors are known to increase the risk for several non-communicable diseases such as type 2 diabetes (T2D), cardiovascular disease and cancer (11–14,16). It is also critical to consider tissue specificity of the epigenome and to test if epigenetic modifications in blood may be used as biomarkers to mimic epigenetic signatures in target tissues for disease.

Adipose tissue is the main energy store in the human body, but also a metabolically active tissue which acts both as an endocrine and an immune organ, and contributes to whole body energy homeostasis (17). Dysfunction of the adipose tissue, e.g. promoted by excessive energy intake, is commonly seen in genetically and environmentally predisposed individuals (18). Adipose tissue gene expression and hormone secretion influence various metabolic phenotypes which, in turn, are associated with human complex traits involved in obesity, T2D and cardiovascular diseases. Epigenetic modifications in adipose tissue may contribute to these phenotypes. Indeed, we recently identified altered gene expression and differential DNA methylation in adipose tissue from subjects with T2D compared with non-diabetic controls (5). We have also shown that regular exercise contributes to extensive transcriptional and DNA methylation changes in human adipose tissue (19,20). Additionally, increased BMI has been associated with increased DNA methylation of *HIF3A* in both human adipose tissue and blood cells (21). However, the potential associations between estimates of obesity or glycemia and the genome-wide DNA methylation pattern in human adipose tissue from non-diabetic subjects have not yet been investigated.

Several studies further point to the importance of epigenetic modifications in the process of aging (3,15,22). We have previously identified age-associated changes in DNA methylation in human skeletal muscle, pancreatic islets and blood cells (3,23–25). More recently, genome-wide, well-powered cross-sectional DNA methylation studies have been performed in leukocytes and whole blood, showing that almost 30 and 15%, respectively, of the analyzed DNA methylation sites were associated with age (26,27). This finding has also been verified in a longitudinal study (28). The age-associated changes in DNA methylation may be influenced by the underlying genetic architecture (24), resulting in both common and tissue-specific alterations (29). However, whether age affect the genome-wide DNA methylation pattern in human adipose tissue and if any of these age-associated epigenetic changes can also be found in blood cells is not known.

The aim of this study was to perform EWAS in human subcutaneous adipose tissue obtained from a discovery cohort of 96 males (male discovery cohort) and in a validation cohort of 94 females (female validation cohort) and relate the genome-wide DNA methylation pattern to three selected known risk factors for common complex diseases (age, BMI and HbA1c). This study design gives us the opportunity to test for both common and gender-specific effects on epigenetic variation. We also investigated the association between the same phenotypes (age, BMI and HbA1c) and genome-wide mRNA expression in adipose tissue from the 96 males. We finally tested if epigenetic variation in blood cells can mirror epigenetic signatures in adipose tissue and potentially be used as epigenetic biomarkers, using adipose tissue and blood cells from a mixed validation cohort (37 males and 67 females) and published data obtained from blood cells.

Results

Analysis of DNA methylation and gene expression in human adipose tissue

To study if known risk factors for common complex diseases, i.e. age, BMI and HbA1c levels (11–14,16), may mediate their effects via epigenetic modifications, we analyzed DNA methylation genome-wide in adipose tissue from 96 males without known disease and with a broad range in age, BMI and HbA1c (male discovery cohort; Table 1). We proceeded to study the impact of age, BMI and HbA1c on DNA methylation levels in adipose tissue from the male discovery cohort using a random effect mixed model, including cohort as the random effect variable and age, BMI and HbA1c as fixed factors. However, we first calculated variance inflation factors (VIFs), which provides information about potential multicollinearity of the studied phenotypes (i.e. age, BMI and HbA1c) (30). Importantly, in the male discovery cohort, all calculated VIFs were close to 1 (1.04–1.18), demonstrating

Table 1. Clinical characteristics of study participants

Characteristic	Male discovery cohort ($n = 96$ males)	Female validation cohort ($n = 94$ females)	Mixed validation cohort ($n = 67$ females, 37 males)
Age (years)	32.4 ± 12.8 (23–80)	29.2 ± 4.2 (21–37)	52 ± 11 (32–83)
BMI (kg/m^2)	25.6 ± 3.7 (17.5–39.0)	27.2 ± 6.7 (18.2–44.9)	27.6 ± 5.2 (18–47)
IFCC HbA1c (mmol/mol)	34 ± 4 (28–46)	31 ± 3 (25–39)	34 ± 3 (22–44)
NGSP HbA1c (%)	5.3 ± 0.3 (4.7–6.4)	5.0 ± 0.3 (4.4–5.7)	5.3 ± 0.3 (4.2–6.2)

Data are expressed as the mean ± SD (range). IFCC HbA1c (mmol/mol), reference value < 50 years: 27–42 and >50 years: 31–46; NGSP HbA1c (%), reference value: 4.0–6.0.

that there are no problems with multicollinearity among the studied phenotypes (age, BMI and HbA1c). Genomic DNA from adipose tissue of these 96 males successfully generated DNA methylation data for 456 800 CpG sites throughout the genome. After correction for multiple testing, we found 62 496 CpG sites significantly associated with one or more of the three phenotypes studied (age, BMI and HbA1c; $q < 0.05$), representing all chromosomes (Supplementary Material, Tables S1–S3). Due to the possible interaction between epigenetic modifications and gene expression (31), we also generated mRNA microarray data from adipose tissue of 94 of the males with available DNA methylation data. Here, mRNA expression data were obtained from a total of 28 499 probe sets representing 22 115 annotated transcripts and 20 246 unique genes. Moreover, after quality control and filtering of probes, genomic DNA from adipose tissue of 94 females included in the female validation cohort (Table 1) (32,33) successfully generated DNA methylation data for 460 973 CpG sites throughout the genome.

Adipose tissue DNA methylation and age

Aging is associated with numerous diseases and it has also been suggested to increase epigenetic variability, including altered levels of DNA methylation (3). This phenomenon includes both common and tissue-specific events (29); however, the specific effect of age on the genome-wide DNA methylation pattern in human adipose tissue is not known. In the male discovery cohort, including 96 males with a range in age between 23 and 80 years, we found that the average DNA methylation level for all 456 800 CpG sites throughout the genome correlated positively with age ($P = 1.1 \times 10^{-5}$, Supplementary Material, Table S4). When dividing the sites based on their relation to the nearest gene (TSS1500, TSS200, 5′UTR, 1st exon, gene body, 3′UTR, intergenic) or in relation to CpG islands (northern shelf, northern shore, CpG island, southern shore, southern shelf, open sea), the average methylation levels were positively and significantly associated with age for all tested regions (Supplementary Material, Table S4). After correction for multiple testing, we found that DNA methylation of 31 567 individual CpG sites in adipose tissue from the male discovery cohort was significantly associated with age ($q < 0.05$), indeed suggesting that the human DNA methylome in adipose tissue changes with age (Supplementary Material, Table S1). Among these sites, 24 514 are annotated to 11 036 unique genes, whereas 7053 of the CpG sites are intergenic. Most of the CpG sites significantly associated with age showed a positive relation between age and methylation level ($n = 28\,605$; 90.6%), whereas only 2962 (9.4%) of the CpG sites showed a negative relation (Fig. 1A). The most significant association between age and DNA methylation was seen for a CpG site upstream of *ELOVL2* (cg21572722, TSS1500, $q = 8.5 \times 10^{-24}$; Fig. 2A, Supplementary Material, Table S1). Altogether, methylation of eight CpG sites annotated to *ELOVL2* correlated significantly with age ($q < 0.05$, Fig. 2A, Supplementary Material, Table S1).

We proceeded to study the genomic distribution of all individual CpG sites significantly associated with age in adipose tissue from the male discovery cohort, either based on their relation to the nearest gene or in relation to CpG islands (Fig. 1C and D and Supplementary Material, Table S5). When comparing the distribution of the significant age-associated CpG sites with the distribution of all analyzed CpG sites across the different genomic regions using χ^2-tests, we found an under-representation of significant CpG sites within TSS1500, TSS200, 5′UTR and intergenic regions, and an over-representation within the 1st exon, gene body and 3′UTR (Supplementary Material, Table S5 and Fig. 1C). We also observed a pronounced effect of age on DNA methylation for CpG sites in regions in relation to CpG islands, with a strong over-representation of significant CpG sites within CpG islands and under-representations in the open sea and shelf regions (Supplementary Material, Table S5 and Fig. 1D).

To test for general and gender-specific effects on epigenetic variation, we next studied the impact of age on DNA methylation in adipose tissue from 94 females (the female validation cohort, Table 1). It should be noted that the span in age was smaller in the female validation cohort (21–37 years) compared with the male discovery cohort (Table 1). Nevertheless, we found DNA methylation of 62 CpG sites to be significantly associated with age in the female validation cohort ($q < 0.05$), where 60 (96.8%) show positive and 2 (3.2%) negative correlations (Supplementary Material, Table S6). In agreement with our result in the male discovery cohort, the most significant association in the female validation cohort was seen for a CpG site upstream of *ELOVL2* (cg16867657, $q = 4.5 \times 10^{-7}$), and methylation in three CpG sites annotated to *ELOVL2* correlated significantly with age in both cohorts, suggesting some common effects of age on methylation in adipose tissue from both males and females ($q < 0.05$, Fig. 2A and B and Supplementary Material, Table S6). Additionally, the majority (42 CpG sites, 70%) of the CpG sites showing a positive correlation with age in the female validation cohort ($q < 0.05$) were also positively correlated with age in the male discovery cohort ($q < 0.05$, Supplementary Material, Table S6). These include CpG sites annotated to *SPATA18*, *PATZ1*, *ANK1*, *NPAS4* and *CADPS2*, genes previously associated with aging (34–38).

To test if epigenetic modifications in blood cells may mirror epigenetic signatures in target tissues for metabolic diseases such as adipose tissue and potentially be used as epigenetic biomarkers, we further compared our results identified in adipose tissue with data from a recent study examining the impact of age on the genome-wide DNA methylation pattern in white blood cells from 421 individuals ranging in age from 14 to 94 years (27). In white blood cells, Johansson *et al.* found that age affected DNA methylation at 137 993 sites, which corresponds to almost one-third of the investigated sites. Interestingly, the

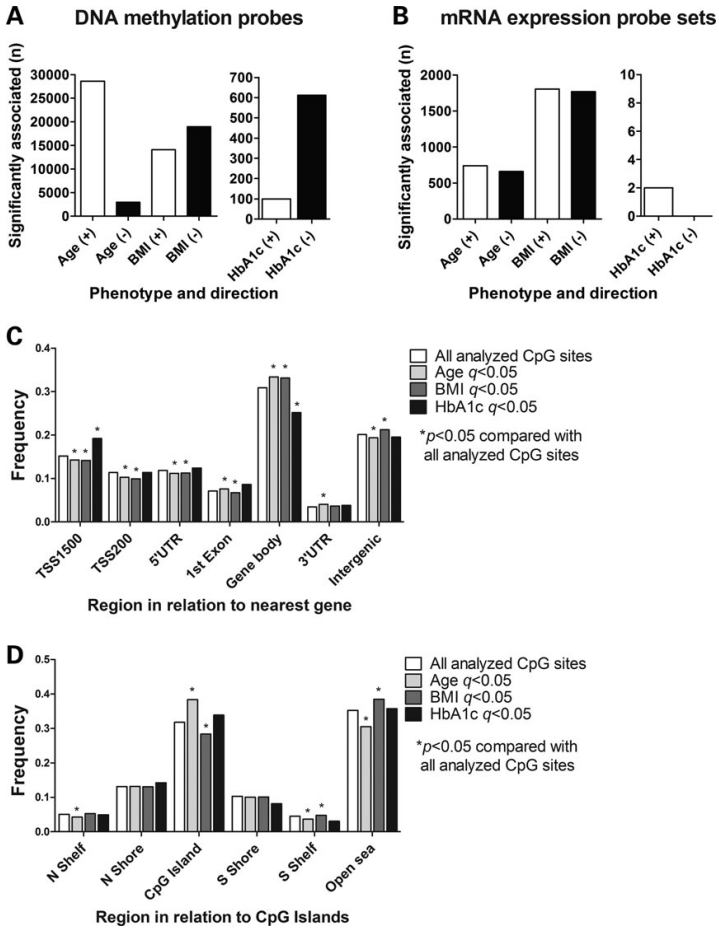


Figure 1. Distribution of significant CpG sites and mRNA expression probe sets in the male discovery cohort. Number and distribution of positive versus negative significant associations between DNA methylation of individual CpG sites (**A**) or mRNA expression probe sets (**B**) and the phenotypes age, BMI and HbA1c. +, positive association; –, negative association. Distribution of CpG sites significantly associated with age, BMI or HbA1c compared with all analyzed CpG sites in relation to gene region (**C**) and CpG island region (**D**).

DNA methylation level of 12 708 of these sites was also significantly associated with age in adipose tissue in the male discovery cohort in the present study, 9897 (78%) in the same direction as in blood (Supplementary Material, Table S7). Moreover, 51 (82%) of the 62 CpG sites significantly associated with age in the female validation cohort were also significantly associated with age in the study of white blood cells (27), all in the same direction in the two studies. All together, the overlap between our two adipose tissue cohorts and blood data by Johansson *et al.* was 37 CpG sites including sites annotated to *ELOVL2*, *SPATA18*, *ANK1*, *NPAS4* and *CADPS2* showing associations between age and DNA methylation (Table 2). Moreover, Steegenga *et al.* (39) have recently summarized data from several studies where the impact of age on DNA methylation in whole blood or purified blood cells was investigated. They presented a list of 14 genes, *ELOVL2*, *FHL2*, *PENK*, *KLF14*, *SST*, *GLRA1*, *TP73*, *GATA4*, *THRB*, *DLX5*, *NEFM*, *TMEM179*, *ATP8A2* and *FOXE3*, displaying age-related changes in DNA methylation based on previous published studies (26,28,40–45). Importantly, all of these previously reported epigenetic biomarkers of aging in blood did also show significant associations with age in adipose tissue from our male discovery cohort (Supplementary Material, Table S1, Table 3 and Fig. 2A, C and D). Also, despite the limited age span in our female validation cohort, four of these genes (*ELOVL2*, *FHL2*, *KLF14* and *GLRA1*) were among the genes significantly associated with age (Supplementary Material, Table S6 and Fig. 2B–D).

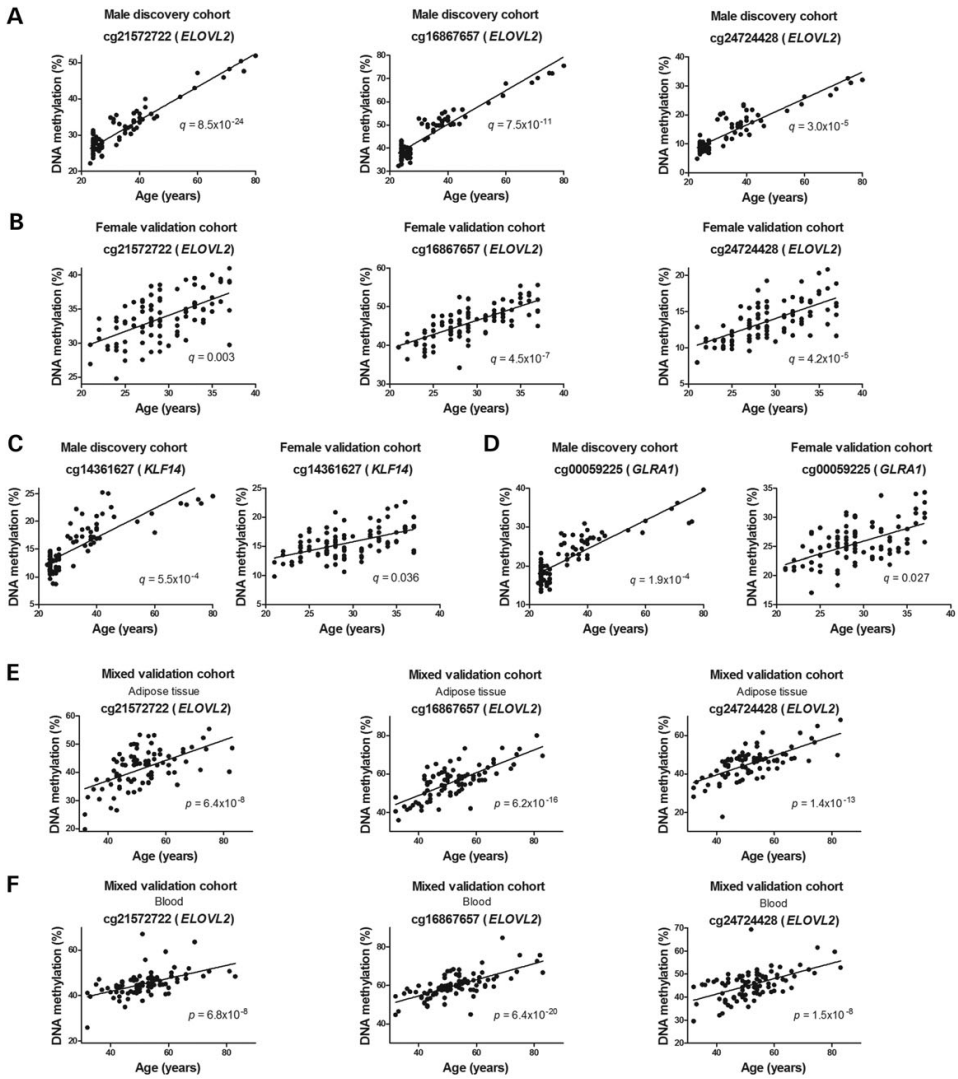


Figure 2. Age correlates with DNA methylation and mRNA expression of specific genes in adipose tissue and blood. DNA methylation at cg21572722, cg16867657 and cg24724428 in *ELOVL2* correlated significantly with age in both the male discovery cohort (**A**) and the female validation cohort (**B**). DNA methylation at CpG sites in *KLF14* (**C**) and *GLRA1* (**D**) correlated significantly with age in both the male discovery cohort and the female validation cohort. DNA methylation at cg21572722 (adipose tissue $n = 81$, blood $n = 83$), cg16867657 ($n = 90$) and cg24724428 (adipose tissue $n = 87$, blood $n = 89$) in *ELOVL2* correlated with age in both adipose tissue (**E**) and blood (**F**) in the mixed validation cohort. DNA methylation in adipose tissue correlated significantly with DNA methylation in blood at cg21572722 ($n = 62$), cg16867657 ($n = 77$) and cg24724428 ($n = 74$) in *ELOVL2* in the mixed validation cohort (**G**). DNA methylation at cg14361627 (adipose tissue $n = 94$, blood $n = 108$) in *KLF14* correlated with age in both adipose tissue (**H**) and blood (**I**) in the mixed validation cohort, and DNA methylation in adipose tissue correlated significantly with DNA methylation in blood ($n = 90$, **J**). For *NOX4* (**K**) both mRNA expression and DNA methylation correlated significantly with age in the male discovery cohort (the most significant CpG site is shown). (**L**) Selected significantly enriched KEGG pathways (FDR adjusted *P*-values < 0.05) of genes that exhibit associations between DNA methylation and age in the male discovery cohort. A complete list of significantly enriched KEGG pathways is presented in Supplementary Material, Table S20.

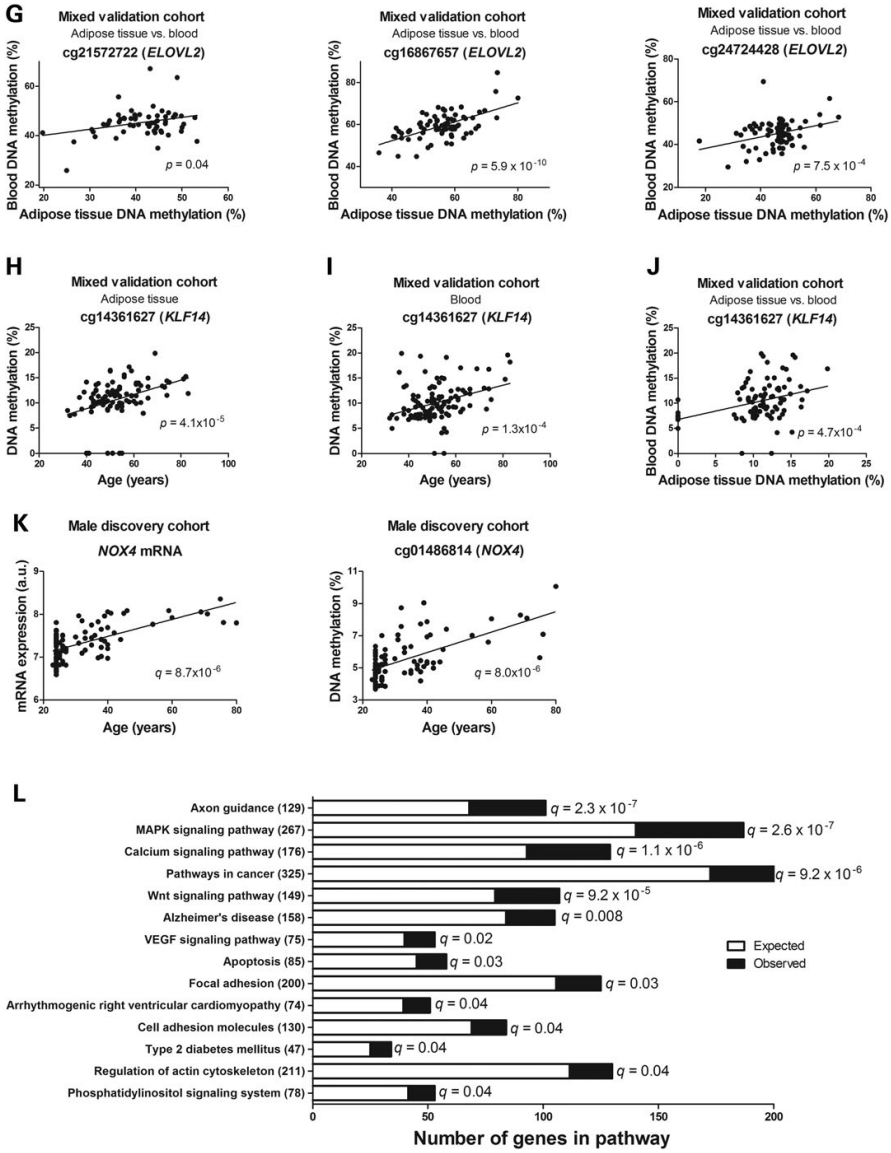


Figure 2. Continued.

We proceeded to test if we could identify age-related epigenetic changes in both adipose tissue and blood cells taken from the same subjects. Here, we used pyrosequencing to analyze DNA methylation of three CpG sites in *ELOVL2* and one CpG site in *KLF14* in both adipose tissue and whole blood cells taken from a mixed validation cohort including 37 males and 67 females with an age span of 32–83 years (Table 1). These CpG sites were selected based on the age-associated changes in DNA methylation in our male discovery cohort (Fig. 2A and C), female validation cohort (Fig. 2B and C) and blood cells in previous published studies

Table 2. Overlap of CpG sites with DNA methylation levels significantly altering with age in adipose tissue from our male discovery cohort and female validation cohort (present study, $q < 0.05$) as well as in white blood cells (27)

Target ID	Chromosome and position	Gene	Relation to gene region	Relation to CpG island	Adipose tissue (96 men)		Adipose tissue (94 women)		Blood (Johansson *et al.*)	
					FDR *q*-value	Regression coefficient	FDR *q*-value	Regression coefficient	*P*-value	Regression coefficient
cg16867657	6:11 044 877	*ELOVL2*	TSS1500	Island	7.5E−11	0.034	4.5E−07	0.038	1.8E−231	0.0051
cg24724428	6:11 044 888	*ELOVL2*	TSS1500	Island	3.0E−05	0.036	4.2E−05	0.043	1.3E−161	0.0043
cg03036557	13:92 050 720	*GPC5*	TSS1500	N_Shore	2.7E−07	0.045	1.7E−04	0.037	7.7E−13	0.0004
cg13516820	9:37 036 494		Intergenic	Island	0.009	0.010	0.001	0.027	2.3E−28	0.0002
cg04880546	12:1 20 868 468		Intergenic	Open Sea	5.6E−04	0.025	0.001	0.039	2.1E−18	0.0005
cg01429039	4:52 918 065	*SPATA18*	Body	Island	0.002	0.019	0.002	0.035	1.1E−23	0.0008
cg23193759	10:71 389 896	*C10orf35*	TSS200	Island	2.1E−06	0.025	0.003	0.037	2.6E−23	0.0008
cg12589298	19:50 828 905	*KCNC3*	Body	Island	0.008	0.022	0.003	0.029	9.5E−13	0.0003
cg23606718	2:1 31 513 927	*FAM123C*	5′UTR;1st Exon	Island	1.8E−06	0.033	0.003	0.030	2.1E−113	0.0026
cg07547549	20:44 658 225	*SLC12A5*	Body	Island	1.0E−13	0.026	0.003	0.033	9.7E−112	0.0032
cg21572722	6:11 044 894	*ELOVL2*	TSS1500	Island	8.5E−24	0.028	0.003	0.026	2.6E−164	0.0025
cg12934382	3:51 741 135	*GRM2*	1st Exon;5′UTR	Island	0.042	0.032	0.004	0.076	1.6E−107	0.0053
cg25090514	5:20 38 743		Intergenic	Island	1.2E−03	0.033	0.008	0.037	2.8E−52	0.0011
cg01908954	15:51 973 764	*SCG3*	5′UTR;1st Exon	Island	9.0E−10	0.026	0.010	0.031	2.0E−57	0.0021
cg00303541	3:51 741 280	*GRM2*	5′UTR	Island	0.042	0.017	0.012	0.045	2.8E−69	0.0037
cg02650266	4:1 47 558 239		Intergenic	Island	5.8E−04	0.030	0.012	0.027	1.6E−93	0.0021
cg25287474	9:14 693 538	*ZDHHC21*	TSS200	Island	0.004	0.024	0.013	0.028	1.1E−29	0.0006
cg27320127	2:47 798 396	*KCNK12*	TSS1500	Island	5.5E−10	0.023	0.019	0.029	1.9E−80	0.0018
cg00590036	6:1 58 957 433	*TMEM181*	TSS200	Island	2.9E−04	0.027	0.022	0.029	9.3E−52	0.0010
cg16547027	18:24 127 588	*KCTD1*	5′UTR;1st Exon	Island	1.8E−04	0.015	0.024	0.026	2.5E−81	0.0016
cg23483765	5:1 56 886 996	*NIPAL4*	TSS200	Island	9.0E−04	0.032	0.026	0.040	7.2E−17	0.0007
cg00059225	5:1 51 304 357	*GLRA1*	1st Exon;5′UTR	Island	1.9E−04	0.022	0.027	0.028	8.1E−99	0.0021
cg11896587	19:39 440 734	*FBXO17*	Body	Island	1.5E−08	0.020	0.027	0.028	6.5E−47	0.0013
cg10947146	8:11 058 710	*XKR6*	1st Exon	Island	0.009	0.022	0.031	0.035	5.6E−14	0.0004
cg19891728	8:41 754 871	*ANK1*	TSS1500	Island	5.0E−06	0.023	0.031	0.033	2.0E−61	0.0012
cg02383785	7:1 27 808 848		Intergenic	Island	3.9E−10	0.023	0.035	0.023	8.3E−98	0.0019
cg14361627	7:1 30 419 116	*KLF14*	TSS1500	Island	5.5E−04	0.022	0.036	0.030	1.5E−142	0.0025
cg01685883	4:1 55 254 497	*DCHS2*	Body	Island	0.026	0.009	0.036	0.035	4.9E−10	0.0004
cg07665387	4:1 34 069 388	*PCDH10*	TSS1500	Island	4.5E−05	0.020	0.036	0.026	1.8E−54	0.0016
cg20899581	6:27 841 230	*HIST1H4L*	1st Exon;TSS1500	Island	1.5E−05	0.026	0.036	0.025	2.3E−16	0.0004
cg26158959	1:2 10 111 162	*SYT14*	TSS1500	N_Shore	9.4E−04	0.020	0.037	0.025	2.5E−52	0.0006
cg14704921	4:53 728 654	*RASL11B*	1st Exon;5′UTR	Island	2.3E−07	0.032	0.039	0.027	2.8E−53	0.0009
cg04768203	11:66 189 188	*NPAS4*	Body	Island	4.4E−04	0.011	0.041	0.025	2.0E−16	0.0003
cg22736354	6:18 122 719	*NHLRC1*	1st Exon	Island	0.041	0.016	0.041	0.029	1.8E−116	0.0021
cg20665157	7:1 21 969 421	*CADPS2*	Body	Open Sea	0.007	0.011	0.048	0.030	1.2E−91	0.0033
cg00664406	3:51 740 875	*GRM2*	TSS1500	Island	0.012	0.021	0.0495	0.035	5.8E−104	0.0037
cg11991058	2:1 68 150 602		Intergenic	Island	0.035	0.013	0.0496	0.028	4.0E−40	0.0012

FDR *q*-value, false discovery rate adjusted *P*-value.

Table 3. DNA methylation and association with age ($q < 0.05$) in adipose tissue from the male discovery cohort among genes displaying multiple probes associated with age in blood or blood cells in previously reported studies

Gene	Identified in	Target ID	*P*-value	FDR *q*-value	Regression coefficient	SEM	Chromosome and position	Relation to gene region	Relation to CpG island
ELOVL2	4, 5, 6, 7, 9, 10	cg21572722	1.9E–29	8.5E–24	0.0278	0.0016	6:11 044 894	TSS1500	CpG island
		cg16867657	1.3E–15	7.5E–11	0.0343	0.0035	6:11 044 877	TSS1500	CpG island
		cg24724428	4.5E–08	3.0E–05	0.0357	0.0060	6:11 044 888	TSS1500	CpG island
		cg16323298	4.2E–04	1.4E–02	0.0134	0.0037	6:11 044 974	TSS1500	CpG island
		cg23642061	1.1E–03	2.5E–02	0.0069	0.0020	6:11 043 631	Gene body	N_Shore
		cg25151806	1.4E–03	2.9E–02	0.0160	0.0048	6:11 045 370	TSS1500	S_Shore
		cg21649660	2.7E–03	4.3E–02	0.0058	0.0019	6:10 995 125	Gene body	Open sea
		cg22143569	7.4E–04	2.0E–02	0.0079	0.0023	6:11 044 541	5′UTR;1st Exon	CpG island
FHL2	4, 5, 6, 7, 9, 10	cg19850931	7.7E–05	5.0E–03	0.0120	0.0029	2:105 993 347	Gene body	Open sea
		cg06907053	2.8E–04	1.1E–02	0.0110	0.0029	2:106 015 869	TSS200;5′UTR;TSS1500	CpG island
		cg26344233	2.4E–03	4.0E–02	0.0061	0.0019	2:106 015 817	TSS200;5′UTR;TSS1500	CpG island
PENK	1, 5, 6, 7, 8, 9, 10	cg16219603	9.4E–07	2.5E–04	0.0179	0.0034	8:57 360 586	TSS1500	CpG island
		cg16419235	4.5E–05	3.5E–03	0.0215	0.0050	8:57 360 613	TSS1500	CpG island
		cg16072688	8.2E–05	5.2E–03	0.0101	0.0024	8:57 360 711	TSS1500	CpG island
		cg19414741	1.5E–06	3.5E–04	0.0123	0.0024	8:57 358 240	Gene body	CpG island
		cg12877723	1.5E–05	1.7E–03	0.0274	0.0060	8:57 358 312	Gene body	CpG island
		cg18742346	6.1E–05	4.3E–03	0.0106	0.0025	8:57 358 625	5′UTR;TSS200	CpG island
		cg04127342	5.0E–04	1.6E–02	0.0160	0.0044	8:57 358 130	Gene body	CpG island
		cg21694941	1.4E–04	7.4E–03	0.0147	0.0037	8:57 358 590	5′UTR;1st Exon	CpG island
KLF14	1, 3, 5, 7, 8, 9, 10	cg08097417	4.4E–15	1.5E–10	0.0177	0.0019	7:130 419 133	TSS1500	CpG island
		cg07955995	1.7E–08	1.5E–05	0.0140	0.0023	7:130 419 159	TSS1500	CpG island
		cg04528819	2.0E–07	8.5E–05	0.0178	0.0032	7:130 418 315	1st Exon	CpG island
		cg14361627	2.8E–06	5.5E–04	0.0221	0.0044	7:130 419 116	TSS1500	CpG island
		cg22285878	3.1E–06	5.8E–04	0.0165	0.0033	7:130 419 173	TSS1500	CpG island
		cg18751682	9.7E–06	1.3E–03	0.0119	0.0025	7:130 419 066	TSS200	CpG island
		cg20426994	5.6E–05	4.0E–03	0.0156	0.0037	7:130 418 324	1st Exon	CpG island
		cg21449170	4.8E–04	1.5E–02	0.0134	0.0037	7:130 419 062	TSS200	CpG island
		cg00094518	1.3E–03	2.8E–02	0.0201	0.0060	7:130 418 549	1st Exon	CpG island
		cg25109431	1.3E–03	2.8E–02	0.0076	0.0023	7:130 419 057	TSS200	CpG island
		cg06533629	1.4E–03	2.9E–02	0.0128	0.0039	7:130 419 370	TSS1500	CpG island
SST	1, 3, 5, 7, 8, 9, 10	cg25478614	6.1E–14	8.4E–10	0.0281	0.0032	3:187 387 866	Gene body	N_Shore
		cg00481951	1.8E–13	2.1E–09	0.0318	0.0037	3:187 387 650	Gene body	N_Shore
		cg14703224	7.9E–05	5.0E–03	0.0106	0.0026	3:187 389 415	TSS1500	S_Shore
		cg16927040	2.7E–04	1.1E–02	0.0127	0.0034	3:187 388 128	1st Exon;5′UTR	CpG island
		cg02164046	2.4E–03	4.0E–02	0.0098	0.0031	3:187 388 148	1st Exon;5′UTR	CpG island
GLRA1	1, 2, 3, 4, 8, 9, 10	cg02071447	9.0E–09	9.3E–06	0.0175	0.0028	5:151 304 542	TSS200	CpG island
		cg08316825	2.7E–05	2.5E–03	0.0177	0.0040	5:151 304 547	TSS200	CpG island
		cg26419265	1.2E–03	2.8E–02	0.0124	0.0037	5:151 303 982	Gene body	N_Shore
		cg00059225	6.1E–07	1.9E–04	0.0224	0.0042	5:151 304 357	1st Exon;5′UTR	CpG island
TP73	1, 2, 3, 4, 8, 9, 10	cg20677901	9.1E–04	2.3E–02	0.0183	0.0053	1:3 568 210	TSS1500	CpG island
		cg01915516	2.8E–03	4.4E–02	–0.0309	0.0100	1:3 568 243	TSS1500	CpG island
		cg07178825	1.2E–05	1.5E–03	0.0214	0.0046	1:3 649 574	Gene body;3′UTR	CpG island
		cg19692322	1.9E–03	3.6E–02	0.0057	0.0018	1:3 621 048	Gene body	N_Shelf

Table continues

Table 3. Continued

Gene	Identified in	Target ID	*P*-value	FDR *q*-value	Regression coefficient	SEM	Chromosome and position	Relation to gene region	Relation to CpG island
		cg03664527	2.4E–03	4.1E–02	0.0058	0.0019	1:3 648 749	Gene body;3′UTR	N_Shore
		cg16554148	3.3E–03	4.8E–02	0.0061	0.0020	1:3 630 882	Gene body	N_Shelf
		cg07382920	1.3E–05	1.5E–03	0.0173	0.0037	1:3 567 646	TSS1500	CpG island
		cg24073122	1.4E–05	1.6E–03	0.0121	0.0026	1:3 567 986	TSS1500	CpG island
		cg01434649	2.3E–03	3.9E–02	0.0193	0.0061	1:3 567 870	TSS1500	CpG island
GATA4	1, 2, 3, 8, 9, 10	cg05930881	2.5E–05	2.4E–03	0.0296	0.0067	8:11 560 540	TSS1500	CpG island
		cg13434842	1.3E–04	7.0E–03	0.0246	0.0062	8:11 567 896	Gene body	S_Shore
		cg24646414	6.3E–04	1.8E–02	0.0152	0.0043	8:11 565 725	5′UTR	CpG island
		cg16803596	2.7E–03	4.3E–02	0.0084	0.0027	8:11 561 498	TSS1500	CpG island
THRB	1, 2, 3, 8, 9, 10	cg04928005	2.8E–06	5.5E–04	0.0128	0.0026	3:24 536 812	TSS1500	CpG island
		cg13790603	5.0E–06	8.1E–04	0.0167	0.0034	3:24 536 478	TSS200;TSS1500	CpG island
		cg24120841	2.6E–05	2.5E–03	0.0096	0.0022	3:24 536 562	TSS1500	CpG island
		cg27622506	4.2E–04	1.4E–02	0.0321	0.0087	3:24 537 160	TSS1500	CpG island
		cg09805010	1.0E–03	2.5E–02	0.0070	0.0021	3:24 536 077	5′UTR	CpG island
		cg27526665	1.1E–03	2.5E–02	0.0194	0.0057	3:24 537 050	TSS1500	CpG island
		cg20985755	1.5E–03	3.1E–02	0.0108	0.0033	3:24 536 474	TSS200;TSS1500	CpG island
		cg07647822	2.5E–03	4.1E–02	0.0240	0.0077	3:24 537 349	TSS1500	CpG island
		cg26139133	3.3E–03	4.8E–02	0.0059	0.0020	3:24 536 765	TSS1500	CpG island
		cg02333852	1.5E–04	7.4E–03	0.0088	0.0022	3:24 536 252	1st Exon;5′UTR	CpG island
DLX5	1, 2, 3, 8, 9, 10	cg08101303	1.5E–03	3.1E–02	0.0114	0.0035	7:96 652 222	Gene body	CpG island
NEFM	1, 3, 7, 8, 9, 10	cg18898125	9.7E–08	5.1E–05	0.0119	0.0021	8:24 770 381	TSS1500	N_Shore
		cg26330518	1.1E–05	1.4E–03	0.0082	0.0017	8:24 770 342	TSS1500	N_Shore
		cg07502389	2.5E–07	9.8E–05	0.0220	0.0039	8:24 771 259	TSS200;TSS1500	CpG island
		cg17078116	1.3E–06	3.1E–04	0.0150	0.0029	8:24 772 344	TSS200;1st Exon	CpG island
		cg02106941	1.5E–05	1.7E–03	0.0150	0.0033	8:24 772 137	TSS1500;1st Exon	CpG island
		cg23290344	5.3E–05	3.9E–03	0.0150	0.0035	8:24 771 466	TSS1500;1st Exon	CpG island
		cg20585869	1.2E–03	2.7E–02	0.0115	0.0034	8:24 772 333	TSS200;1st Exon	CpG island
		cg22562942	2.1E–03	3.8E–02	0.0116	0.0037	8:24 772 309	TSS200;1st Exon	CpG island
		cg18267374	1.2E–03	2.7E–02	0.0156	0.0047	8:24 771 273	TSS1500;5′UTR;1st Exon	CpG island
		cg26980244	2.2E–03	3.9E–02	0.0105	0.0033	8:24 772 513	5′UTR;1st Exon;Gene body	CpG island
TMEM179	1, 3, 7, 8, 9, 10	cg00107187	9.8E–10	1.7E–06	0.0151	0.0022	14:1 05 070 998	1st Exon	CpG island
		cg10281977	2.2E–07	9.1E–05	0.0111	0.0020	14:1 05 070 864	1st Exon	CpG island
ATP8A2	1, 2, 3, 4, 8, 9, 10	cg13096859	1.6E–08	1.4E–05	0.0150	0.0024	13:26 310 861	Gene body	Open sea
		cg00255889	1.3E–05	1.5E–03	0.0168	0.0036	13:26 586 288	Gene body	CpG island
		cg18836626	4.2E–05	3.3E–03	0.0104	0.0024	13:26 042 746	Gene body	CpG island
		cg07973479	1.6E–04	7.7E–03	0.0108	0.0027	13:25 945 255	TSS1500	N_Shore
		cg13537353	2.6E–04	1.1E–02	0.0196	0.0051	13:26 586 651	Gene body	CpG island
		cg02057782	4.1E–04	1.4E–02	0.0118	0.0032	13:26 042 598	Gene body	N_Shore
		cg25516742	2.1E–03	3.7E–02	0.0081	0.0025	13:26 503 010	Gene body	Open sea
FOXE3	1, 3, 4, 7, 8, 9, 10	cg22796507	1.5E–11	6.6E–08	0.0144	0.0019	1:47 882 739	1st Exon	CpG island
		cg18815943	2.5E–09	3.5E–06	0.0177	0.0027	1:47 882 314	1st Exon	CpG island
		cg01281911	1.1E–07	5.5E–05	0.0148	0.0026	1:47 882 686	1st Exon	CpG island
		cg19809499	4.0E–04	1.4E–02	0.0130	0.0035	1:47 882 265	1st Exon	CpG island

1, Teschendorff *et al.* (27 K) (44); 2, Raykan *et al.* (27 K) (43); 3, Bell *et al.* (27 K) (40); 4, Heyn *et al.* (450 K) (42), 5, Hannum *et al.* (450 K) (26); 6, Garagnani *et al.* (450 K) (41); 7, Florath *et al.* (450 K) (28); 8, Xu *et al.* (27 K) (45); 9, Steegenga *et al.* (450 K) (39); 10, Johansson *et al.* (450 K) (27); FDR *q*-value, false discovery rate adjusted *P*-value; SEM, standard error of the mean.

(Table 3). Importantly, increased age was significantly associated with increased DNA methylation in these three *ELOVL2* sites and the *KLF14* site in both adipose tissue and blood cells taken from the mixed validation cohort, and the DNA methylation in adipose tissue correlated significantly with the methylation in blood (Fig. 2E–J). Together these data show that age-associated methylation changes found in blood cells can mirror epigenetic signatures in target tissues such as adipose tissue and potentially be used as epigenetic biomarkers to predict susceptibility and progression of disease.

We therefore proceeded to test if age-associated methylation differences also can be identified in diseased subjects, i.e. in patients with T2D. We recently published a case–control study identifying 15 627 CpG sites with differential DNA methylation in adipose tissue from subjects with T2D compared with non-diabetic age-matched controls (5). As aging is a known risk factor for T2D, we further investigated the overlap between the 31 567 CpG sites showing significant association between DNA methylation and age in adipose tissue from our discovery cohort of 96 non-diabetic males in the present study (Supplementary Material, Table S1) and the 15 627 sites showing differential DNA methylation in adipose tissue from subjects with T2D compared with controls in our previous study (5). Notably, DNA methylation of 1278 CpG sites was both significantly associated with age in the 96 non-diabetic males and displayed differential DNA methylation between T2D subjects and age-matched controls. Importantly, DNA methylation of as many as 90% of these CpG sites changed in the same direction due to increasing age or T2D (Supplementary Material, Table S8). These include CpG sites annotated to *IRS1* and *KCNQ1*, genes previously implicated in the pathogenesis of T2D (46,47), as well as to *TMEM17*, an age-associated blood-based epigenetic biomarker (Table 3). In total, DNA methylation of 188 CpG sites was significantly associated with age in both adipose tissue (male discovery cohort) and blood (in the study by Johansson *et al.*) and displayed differential DNA methylation between T2D subjects and controls in the same direction (Supplementary Material, Table S7). These data support the use of blood-based epigenetic biomarkers to foresee epigenetic changes that take place in target tissues of diseased subjects, i.e. with T2D.

Adipose tissue mRNA expression and age

We proceeded to test if age was associated with altered gene expression in human adipose tissue. Interestingly, we also found a striking effect of age on adipose tissue mRNA expression levels in the male discovery cohort, with expression of 1400 probe sets of which 1130 are annotated to 1084 unique genes significantly associated with age ($q < 0.05$, Supplementary Material, Table S9). The number of positive and negative correlations between mRNA expression and age is shown in Figure 1B. The most significant correlations between age and mRNA expression were observed for *NOX4* (positive correlation, $q = 8.7 \times 10^{-6}$, Fig. 2K) and *PLG* (negative correlation, $q = 8.7 \times 10^{-6}$, Supplementary Material, Table S9). As epigenetic modifications are known to regulate tissue-specific gene expression (31), we further tested if genes with significant correlations between mRNA expression and age also showed age-related changes in DNA methylation. Indeed, in 1050 of the 1084 genes with age-associated changes in mRNA expression, DNA methylation was also associated with age in the male discovery cohort (Supplementary Material, Table S10) when using CpG sites within the *cis* distance 500 kb upstream and 100 kb downstream of each gene. These include *NOX4* (Fig. 2K), *PLG*, *ETS2*, *CCR2* and *CXCR2* (Supplementary Material, Table S10).

Adipose tissue DNA methylation and BMI

BMI is a simple measurement of body size and widely used for population based studies, as well as for diagnosis of overweight (BMI > 25 kg/m^2) and obesity (BMI > 30 kg/m^2). Increased BMI is associated with the development of several common diseases (12,13). There are metabolic differences observed in adipose tissue from obese compared with lean subjects (48) and these changes may partly be due to epigenetic modifications. In the adipose tissue from the male discovery cohort with a range in BMI between 17.5 and 39.0 kg/m^2, the average DNA methylation level for all 456 800 CpG sites throughout the genome did not correlate significantly with BMI ($P = 0.3$, Supplementary Material, Table S4). However, we found that DNA methylation of 33 058 individual CpG sites was significantly associated with BMI in the male discovery cohort ($q < 0.05$), of which 24 939 sites are annotated to 12 325 unique genes and 8119 sites are intergenic (Supplementary Material, Table S2). In contrast to our identified associations between adipose tissue DNA methylation and age, there were slightly more negative ($n = 18\,972$, 57.4%) than positive ($n = 14\,086$, 42.6%) significant associations between DNA methylation and BMI (Fig. 1A). Among annotated genes, the most significant correlation between BMI and DNA methylation was observed for a CpG site in the promoter of *CCRL2* (cg18599081, negative correlation, $q = 9.0 \times 10^{-6}$; Fig. 3A). Among all CpG sites significantly associated with BMI, we observed an over-representation within the gene body and intergenic regions, and less significant CpG sites compared with the distribution on the array in the region surrounding transcription start, i.e. TSS1500, TSS200, 5′UTR and 1st exon (Supplementary Material, Table S5 and Fig. 1C). In contrast to what was seen for associations between age and DNA methylation, CpG sites with DNA methylation associated with BMI were under-represented within CpG islands and over-represented within the open sea and southern shelf (Supplementary Material, Table S5 and Fig. 1D).

We next investigated the impact of BMI on DNA methylation in adipose tissue from the female validation cohort including 94 females with a range in BMI between 18.2 and 44.9 kg/m^2. Here, we found that DNA methylation of 39 533 CpG sites was significantly associated with BMI ($q < 0.05$), of which 30 507 sites are annotated to 11 766 unique genes and 9026 sites are intergenic (Supplementary Material, Table S11). These include 27 809 (70%) positive and 11 724 (30%) negative correlations between DNA methylation and BMI. The most significant correlation was observed for a CpG site in the body of *PLEC1* (negative correlation; $q = 1.8 \times 10^{-12}$, Fig. 3B). This site was also negatively correlated with BMI in the male discovery cohort (Fig. 3B). The overlap between the male discovery cohort and the female validation cohort was 4979 CpG sites, 2756 sites with positive and 2223 sites with negative correlations between BMI and DNA methylation (Supplementary Material, Table S12). These include sites annotated to genes previously linked to obesity, T2D and/or fat metabolism, e.g. *FTO*, *TCF7L2*, *FASN*, *IRS1*, *IRS2*, *MTCH2* and *PPARGC1B*. Many of the most significant sites in the male discovery cohort could be validated in the 94 females ($q < 0.05$), including cg18599081 in *CCRL2* (Fig. 3A), cg12492380 in *MYH16*, cg27115863 (intergenic) and cg11151251 in *DCAF5* among the 10 most significant sites. Importantly, since the discovery cohort only includes males and the validation cohort only females, the overlapping sites are most likely gender-unspecific, while the sites correlating only in males or females might represent gender-specific associations between BMI and adipose tissue DNA methylation. Impressively, we were able to replicate the strong, gender-unspecific association between BMI and DNA

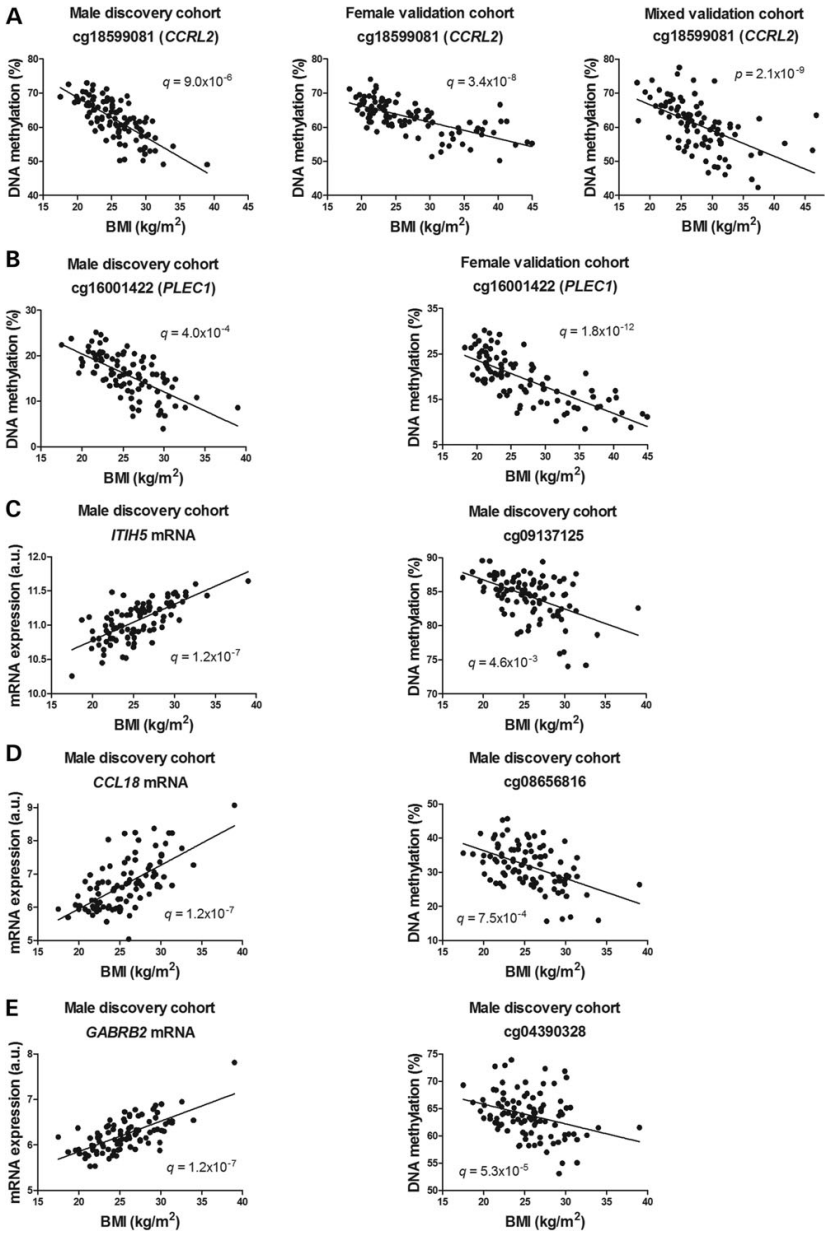


Figure 3. Correlations between BMI and DNA methylation and BMI and mRNA expression in human adipose tissue. DNA methylation at cg18599081 in *CCRL2* (**A**) correlated significantly with BMI in the male discovery cohort, the female validation cohort as well as the mixed validation cohort ($n = 91$). DNA methylation at cg16001422 in *PLEC1* (**B**) correlated significantly with BMI in both the male discovery cohort and the female validation cohort. For *ITIH5* (**C**), *CCL18* (**D**) and *GABRB2* (**E**), both DNA methylation and mRNA expression correlated with BMI in the male discovery cohort (the most significant CpG site within the *cis* distance 500 kb upstream and 100 kb downstream of each gene is shown). (**F**) Selected significantly enriched KEGG pathways (FDR adjusted P-values < 0.05) of genes that exhibit correlations between DNA methylation and BMI in the male discovery cohort. A complete list of significantly enriched KEGG pathways is presented in Supplementary Material, Table S21.

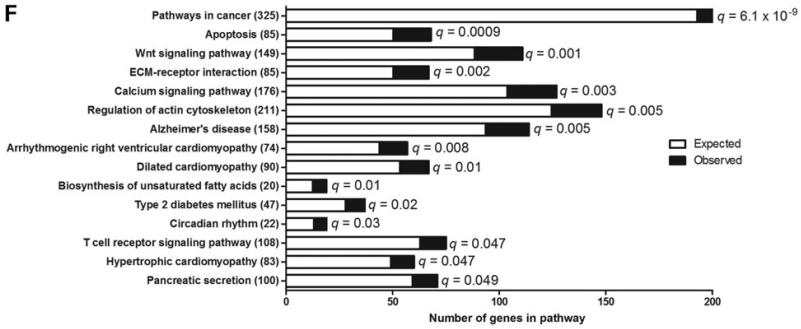


Figure 3. Continued.

Table 4. DNA methylation in *HIF3A* in relation to BMI, in adipose tissue from the male discovery cohort and the female validation cohort

Target ID	Chromosome and position	Gene	Relation to gene region	Relation to CpG island	Adipose tissue Male discovery cohort		Adipose tissue Female validation cohort	
					FDR *q*-value	Regression coefficient	FDR *q*-value	Regression coefficient
cg01552731	19:46 806 907	*HIF3A*	1st Exon;5′UTR;Body	N_Shore	0.039	0.045	0.0007	0.033
cg23548163	19:46 807 119	*HIF3A*	5′UTR;Body	Island	ns (0.59)		0.015	0.019
cg16672562[a]	19:46 801 672	*HIF3A*	5′UTR;Body;1st Exon	S_Shore	ns (0.77)		0.019	0.029
cg12068280	19:46 804 528	*HIF3A*	5′UTR;Body	N_Shelf	ns (0.30)		0.028	0.014
cg22891070[a]	19:46 801 642	*HIF3A*	Body;TSS200	S_Shore	ns (0.83)		0.031	0.027
cg07684068	19:46 807 660	*HIF3A*	Body	S_Shore	0.026	0.047	0.032	0.019
cg27146050[a]	19:46 801 557	*HIF3A*	Body;TSS200	S_Shore	ns (0.90)		0.042	0.016

The seven CpG sites where DNA methylation correlated significantly with BMI ($q < 0.05$) in the female validation cohort are presented.
[a]CpG sites reported to be significantly associated with BMI in the study by Dick *et al.* (21). FDR *q*-value, false discovery rate adjusted *P*-value; ns, non-significant *q*-values (>0.05).

methylation of cg18599081 in *CCRL2* in adipose tissue from the mixed validation cohort using pyrosequencing ($P = 2.1 \times 10^{-9}$, Fig. 3A).

We proceeded to compare our results with data from a recently published study, which identified positive correlations between BMI and DNA methylation at three CpG sites in *HIF3A* in adipose tissue (only females) and blood (21). In adipose tissue from our female validation cohort, we observed that DNA methylation of seven CpG sites annotated to *HIF3A*, including the same three CpG sites as reported by Dick *et al.*, were significantly associated with BMI ($q < 0.05$, Supplementary Material, Table S11, Table 4). In contrast, the 96 males of our discovery cohort did not show any significant association between DNA methylation of these *HIF3A* CpG sites and BMI ($q > 0.05$, Table 4). However, DNA methylation of two other sites within the *HIF3A* gene was significantly associated with BMI in our male discovery cohort ($q < 0.05$, Supplementary Material, Table S2, Table 4). These data support both gender-specific and general effects of BMI on the DNA methylation pattern in human adipose tissue.

As obesity, as well as age, is a known risk factor for T2D, we investigated how many of the CpG sites significantly associated with BMI in adipose tissue from non-diabetic subjects in the present study that also show differential DNA methylation in adipose tissue from subjects with T2D compared with non-diabetic controls in our recently published case–control study (5). We found that BMI was significantly associated with DNA methylation of 988 and 3425 CpG sites in the male discovery and female validation cohort, respectively, that also displayed differential DNA methylation between subjects with T2D and controls (Supplementary Material, Tables S13 and S14). The majority (60% in the male discovery cohort and 99% in the female validation cohort) of CpG sites that exhibit BMI associated changes in DNA methylation in non-diabetic subjects changed in the same direction in subjects with T2D compared with controls (Supplementary Material, Tables S13 and S14). These include CpG sites annotated to genes previously linked to T2D, obesity and/or energy metabolism such as *PC*, *FOXO1* and *HSF1* in both cohorts and *IGF2*, *VEGFA*, *IRS1*, *IL1RN*, *IGF1R*, *ELOVL6* and *KCNQ1* in the female validation cohort.

Adipose tissue mRNA expression and BMI

We further studied the association between BMI and gene expression in adipose tissue from the male discovery cohort. Also BMI was associated with transcriptional changes in adipose tissue, with significant associations to 3575 probe sets, of which 3104 are annotated to 2936 unique genes (Supplementary Material, Table S15). The most significant correlations between BMI and mRNA expression were observed for *ITIH5*, *CCL18* and *GABRB2* (positive correlations, $q = 1.2 \times 10^{-7}$, Fig. 3C–E), and *SNORD115-1* (negative correlation, $q = 1.2 \times 10^{-7}$, Supplementary Material, Table S15). Furthermore, 2825 of the 2936 genes with significant associations between BMI and mRNA expression ($q < 0.05$) had one or more CpG sites within the *cis* distance 500 kb upstream

and 100 kb downstream of each gene significantly associated with BMI (Supplementary Material, Table S16). These include *ITIH5*, *CCL18*, *GABRB2*, *FTO*, *MTCH2*, *IRS1* and *SPP1* (*OPN*) (Fig. 3C–E).

Adipose tissue DNA methylation and HbA1c

Glycated hemoglobin (HbA1c) is a long-term measure of average blood glucose levels and elevated HbA1c levels are associated with an increased risk of T2D, cardiovascular disease and cancer (11,14,16). Additionally, glucose-induced epigenetic changes have been suggested to explain the so called metabolic memory, which may increase the risk for cardiovascular disease and T2D (49–52). As adipose tissue has a role in whole body glucose homeostasis, we investigated the association between adipose tissue DNA methylation and HbA1c as a continuous variable. In the male discovery cohort, including 96 non-diabetic males with a range in HbA1c between 28 and 46 mmol/mol (representing 4.7–6.4%), we found that the average DNA methylation level for all 456 800 CpG sites throughout the genome correlated negatively with HbA1c ($P = 0.025$, Supplementary Material, Table S4). Furthermore, we found that DNA methylation of 711 individual CpG sites was significantly associated with HbA1c ($q < 0.05$), of which 541 are annotated to 583 unique genes and 170 CpG sites are intergenic (Supplementary Material, Table S3). Among these CpG sites, 99 (14%) showed positive and 612 (86%) showed negative correlations between adipose tissue DNA methylation and HbA1c (Fig. 1A). The most significant correlation between HbA1c and adipose tissue DNA methylation was seen for a CpG site upstream (TSS1500) of *ANKRD11* (negative correlation; Fig. 4A and Supplementary Material, Table S3), located in a CpG

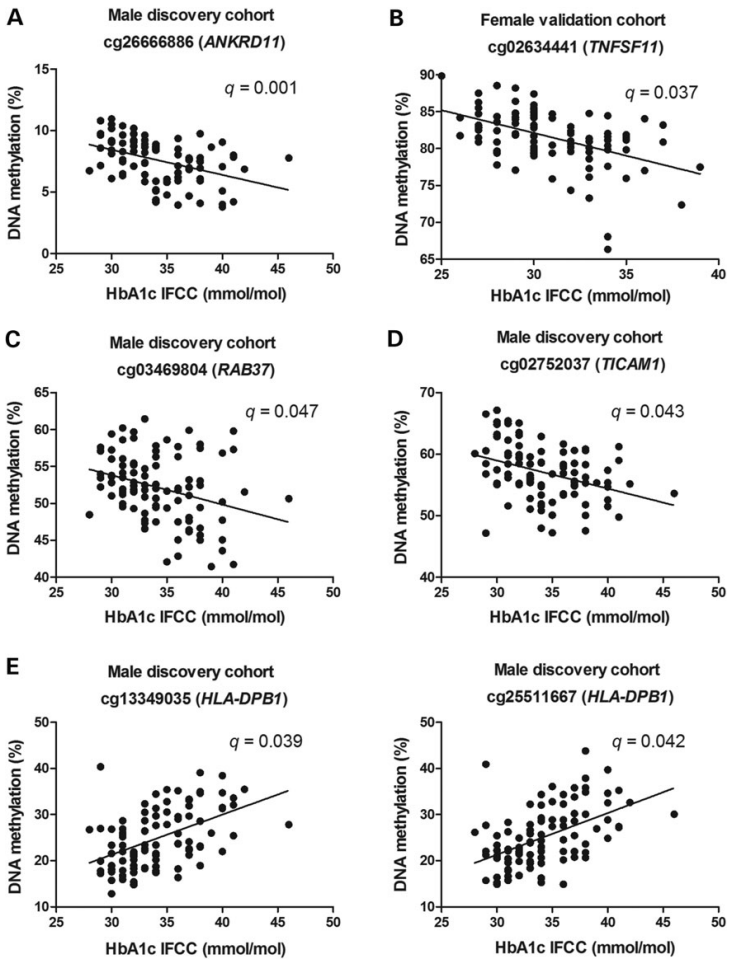


Figure 4. Correlations between HbA1c and DNA methylation in human adipose tissue. DNA methylation at CpG sites in *ANKRD11* (**A**), *TNFSF11* (**B**), *RAB37* (**C**), *TICAM1* (**D**) and *HLA-DPB1* (**E**) correlated significantly with HbA1c.

island shore. Moreover, CpG sites with DNA methylation significantly associated with HbA1c were over-represented within the TSS1500 and under-represented within the gene body regions, whereas no significant difference was observed for the distribution in regions in relation to CpG islands (Supplementary Material, Table S5, Fig. 1C and D).

In our female validation cohort, consisting of 94 non-diabetic females with a span in HbA1c between 25 and 39 mmol/mol (representing 4.4–5.7%), we identified seven CpG sites with DNA methylation significantly associated with HbA1c ($q < 0.05$), two with a positive and five with a negative coefficient (Supplementary Material, Table S17). The strongest correlation was observed for a CpG site upstream *TNFSF11* ($q = 0.04$, Fig. 4B). None of these seven sites were significantly associated with HbA1c in the male discovery cohort.

Elevated HbA1c levels within the non-diabetic interval predict future risk of T2D (16). We therefore tested if CpG sites significantly associated with HbA1c levels in the present study were also found to be differentially methylated in adipose tissue from subjects with T2D compared with controls in our previous case–control study (5). Indeed, we found 30 among the 711 CpG sites significantly associated with HbA1c in the male discovery cohort also to have differential DNA methylation levels in subjects with T2D compared with controls. The majority (28 sites; 93%) of the CpG sites that exhibit differential DNA methylation due to increased HbA1c in non-diabetic subjects changed in the same direction in subjects with T2D compared with non-diabetic controls (Supplementary Material, Table S18). These include CpG sites annotated to genes previously linked to T1D or T2D such as *RAB37*, *TICAM1* and *HLA-DPB1* (53–55) (Fig. 4C–E).

Adipose tissue mRNA expression and HbA1c

We found significant associations between mRNA expression of two transcripts, annotated to LOC100288814 and *CLLU1*, respectively, and HbA1c levels in adipose tissue from the discovery cohort, both with positive correlations (Fig. 1B). However, neither LOC100288814 nor *CLLU1* had CpG sites with DNA methylation significantly associated with HbA1c levels.

Overlap between associations of DNA methylation and age, BMI and HbA1c

It has been suggested that certain regions or positions in the genome are more prone to epigenetic variation (56). Based on this, we investigated the overlap of CpG sites with DNA methylation significantly associated in the same direction with more than one of the phenotypes examined in the male discovery cohort. Most overlap was found between age and BMI, namely 1334 CpG sites in the same direction (Supplementary Material, Table S19). The overlap between age and HbA1c included 2 CpG sites, and between BMI and HbA1c, we found 12 overlapping CpG sites in the same direction. The overlap between associations of DNA methylation and the different phenotypes is hence modest, but likely a result of the statistical model where all three phenotypes are included and thereby adjusted for. However, the overlaps we do detect thereby represent independent effects of the different phenotypes on the DNA methylome.

Pathway analysis

To gain further biological relevance of the significant associations between DNA methylation and the studied phenotypes, we performed KEGG pathway analyses with WebGestalt (http://www.webgestalt.org). We included genes with one or more CpG site(s) annotated to the gene significantly associated with respective studied phenotypic trait ($q < 0.05$). For associations between DNA methylation and age, 31 KEGG pathways involved in, for example, cancer, signal transduction, cardiovascular disease and T2D were significantly enriched in the male discovery cohort (Supplementary Material, Table S20 and Fig. 2L). Among these 31 pathways, only one was also enriched in the female validation cohort (Neuroactive ligand-receptor interaction pathway). For associations between DNA methylation and BMI, 47 KEGG pathways involved in, for example, cancer, signal transduction, cardiovascular disease, T2D and inflammation were significantly enriched in the male discovery cohort (Supplementary Material, Table S21 and Fig. 3F). Among these 47 pathways, 41 (87%) were also enriched in the female validation cohort (Supplementary Material, Table S21). For associations between DNA methylation and HbA1c, no KEGG pathways were significantly enriched.

For transcripts with positive associations between mRNA expression and age, three KEGG pathways involved in cardiovascular disease were significantly enriched (Supplementary Material, Table S22). Pathway analysis of transcripts with mRNA expression negatively associated with age revealed three significant pathways, of which two are involved in chemokine signaling and cytokine–cytokine receptor interaction (Supplementary Material, Table S22).

For positive associations between mRNA expression and BMI, 29 KEGG pathways involved in, for example, the immune system and glycan biosynthesis were significantly enriched (Supplementary Material, Table S23). Pathway analysis of transcripts with mRNA expression negatively associated with BMI revealed six significant pathways involved in, for example, fatty acid metabolism, glucose metabolism, amino acid metabolism and translation (Supplementary Material, Table S23).

Due to uneven genomic distribution of CpG sites and CpG islands and the design of the Illumina 450k array, conclusions based only on pathway analysis could be severely biased (57). However, the pathway results in our study in many cases mirror the results obtained from the analysis of individual genes; hence, we use this additional method in support of our findings and to share lights on metabolic pathways and biological function.

Correlations between DNA methylation of CpG sites significantly associated with age, BMI or HbA1c and mRNA expression in human adipose tissue

DNA methylation is known to regulate gene expression and depending on the genomic location of a CpG site, methylation may be either negatively or positively associated with transcriptional activity (31). To study the direct correlations between DNA methylation and mRNA expression, we performed Spearman's correlations between DNA methylation of individual CpG sites significantly associated with age, BMI or HbA1c (Supplementary Material, Tables S1–S3) and expression of nearby mRNA probe sets in adipose tissue from 94 individuals in the male discovery cohort with genome-wide data available for both DNA methylation and mRNA expression. In this analysis, we included mRNA probe sets within a *cis* distance from the significant CpG sites, i.e. CpG sites within a distance 500 kb upstream and 100 kb downstream of mRNA probe sets. Based on these inclusion criteria, 794 515 CpG–mRNA combinations, including 61 932 unique CpG sites and 28 041 mRNA probe sets, were included in the correlation analysis. After correction for multiple testing, 199 450 of these CpG–mRNA combinations showed a significant correlation ($q < 0.05$) between DNA methylation and mRNA expression.

We further separated the significant correlations between DNA methylation and mRNA expression based on what phenotype the included CpG sites were associated with. For CpG sites associated with age, we found 33 000 positive and 43 825 negative CpG–mRNA correlations including 17 710 unique CpG sites and 17 056 unique mRNA probes. Thus, 56% of all CpG sites significantly associated with age also show a correlation with one or more mRNA expression probe sets (Supplementary Material, Table S24). For example, the CpG site most significantly associated with age, cg21572722 in *ELOVL2* and several other CpG sites in *ELOVL2* showed significant correlations to mRNA expression (Supplementary Material, Table S24). Some of these correlations are presented in Figure 5A. Additionally, among the 33 058 CpG sites significantly associated with BMI, 23 361 unique sites (71%) were also significantly associated with expression of one or more mRNA probe sets, giving rise to 62 134 positive and 66 679 negative CpG–mRNA correlation combinations (Supplementary Material, Table S25). Included are, for example, cg18599081 in *CCRL2* (Fig. 5B) and cg01552731 (four CpG–mRNA combinations) and cg07684068 (nine CpG–mRNA combinations) in *HIF3A*. For HbA1c, we found 891 positive and 1095 negative correlations between DNA methylation and mRNA expression, including 397 unique CpG sites (Supplementary Material, Table S26), i.e. 56% of all the CpG sites significantly associated with HbA1c.

Enzymes regulating DNA methylation

DNA methylation in mammalian cells is carried out by DNA methyltransferases (DNMTs) (1), making this family of enzymes inevitable for a functional epigenome. We therefore investigated if age, BMI or HbA1c were associated with altered DNA methylation and/or mRNA expression of the genes encoding these enzymes in human adipose tissue. The Infinium HumanMethylation450 BeadChip array analyzes DNA methylation of 134 CpG sites annotated to *DNMT1*, *DNMT3A*, *DNMT3B* or *DNMT3L*. In the male discovery cohort, we found DNA methylation of 11 CpG sites annotated to these genes significantly associated with age (all with a positive coefficient) and another 10 sites associated with BMI (6 negatively and 4 positively associated; $q < 0.05$), (Supplementary Material, Tables S1 and S2). We also found that age has a negative impact on *DNMT3A* mRNA levels in the male discovery cohort ($q = 0.04$, Supplementary Material, Table S9). In the female validation cohort, we found DNA methylation of 21 CpG sites annotated to genes encoding DNMTs significantly associated with BMI (6 negatively and 15 positively associated; $q < 0.05$; Supplementary Material, Table S11).

We finally investigated the TET enzymes, which have a role in DNA demethylation by creating hydroxymethylation (58). Of the 63 CpG sites analyzed in *TET1*, *TET2* and *TET3* in the male discovery cohort, we found four sites with DNA methylation positively associated with age as well as six sites positively and one negatively associated with BMI ($q < 0.05$; Supplementary Material, Tables S1 and S2). Furthermore, *TET3* mRNA expression showed a positive association with BMI ($q = 0.02$, Supplementary Material, Table S15). In the female validation cohort, we found four CpG sites annotated to genes encoding TET enzymes with DNA methylation positively associated with BMI ($q < 0.05$; Supplementary Material, Table S11).

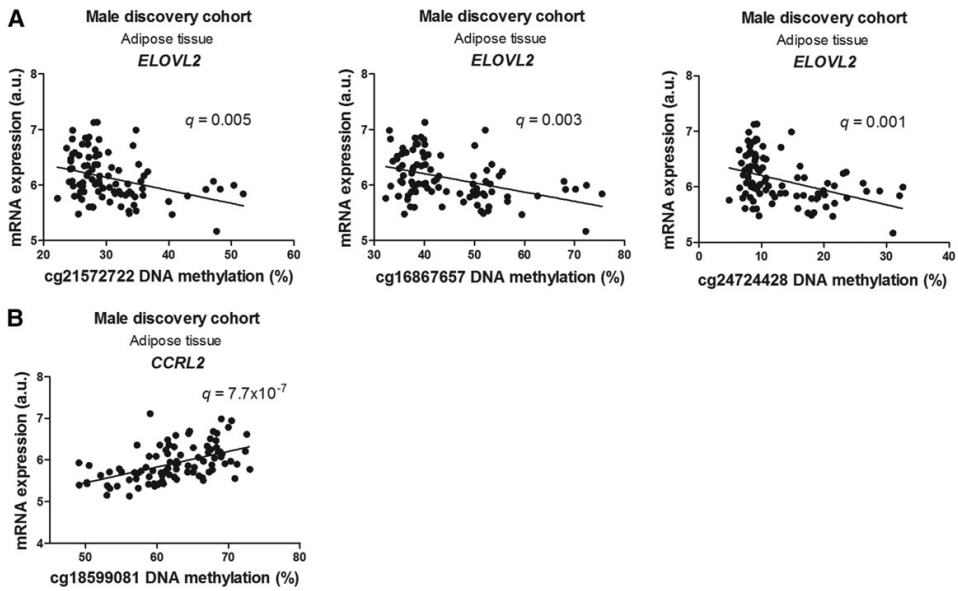


Figure 5. Correlations between DNA methylation of CpG sites significantly associated with age or BMI and mRNA expression in human adipose tissue. DNA methylation at cg21572722, cg16867657 and cg24724428 in *ELOVL2* correlated significantly with mRNA expression of *ELOVL2* (probe set 8123920) in adipose tissue from the male discovery cohort (**A**). DNA methylation at cg18599081 in *CCRL2* correlated significantly with mRNA expression of *CCRL2* (probe set 8079407) in adipose tissue from the male discovery cohort (**B**).

Discussion

In this study, we have shown that three known risk factors for common diseases, i.e. age, BMI and HbA1c play important roles in determining the pattern of DNA methylation and mRNA expression in human adipose tissue. These changes take place in genes known to contribute to development of age- and obesity-related diseases such as T2D, cardiovascular disease and cancer. We also demonstrate for the first time that age-associated epigenetic variation in blood can mirror epigenetic signatures in adipose tissue and potentially be used as epigenetic biomarkers for metabolic diseases.

Identification of subjects with a high risk of developing metabolic disease is a criterion for disease prevention. Genetic and environmental factors have been shown to predict T2D and obesity (13). However, their capacity to predict disease is still suboptimal and there is a need for new biomarkers with a high capacity of predicting metabolic disease. These may potentially include epigenetic biomarkers. Importantly, the epigenome is dynamic and changes due to environmental exposures (3,19,24,25,59–62) but once epigenetic modifications are introduced they may be stable and inherited through cell divisions (63,64), making epigenetics a potentially important pathogenic mechanism in complex diseases. However, it is important to be aware of the tissue-specific nature of the epigenome. Although the nucleotide sequence is identical in most human cells, the epigenetic pattern is highly cell specific as it contributes to the diverse phenotype and the expression pattern seen in different cell types (6,51,65). In most large well-phenotyped cohorts, blood is often the only source of biological material available, and so far little is known about the correlation between DNA methylation in target tissues such as adipose tissue and more accessible cell types such as blood cells. Additionally, blood-based biomarkers have an important clinical relevance since they are easy to analyze in patients and high-risk populations. In this study, many epigenetic biomarkers of aging in blood, i.e. *ELOVL2*, *FHL2*, *KLF14* and *GLRA1*, also showed significant correlations between adipose tissue DNA methylation and age. Notably, our data demonstrate that epigenetic biomarkers in blood can mirror age-related epigenetic signatures in biologically relevant target tissues such as adipose tissue.

Many of the genes associated with age in our study have functions related to the aging process and/or are previously reported biomarkers of aging in blood. For example, DNA methylation of *ELOVL2*, encoding fatty acid elongase 2, has previously been shown to increase with age in blood cells (41) and in our study the most significant association between age and DNA methylation was found upstream of *ELOVL2*. Importantly, we found significant correlations between methylation in several CpG sites annotated to *ELOVL2* and age in adipose tissue from three different cohorts as well as in whole blood cells, supporting the use of blood-based epigenetic biomarkers to mirror some epigenetic signatures in target tissues. Also, *KFL14* belongs to the genes showing increased DNA methylation in both human adipose tissue and blood cells with increased age. *KFL14* encodes Kruppel-like factor 14 which has been suggested to be a master regulator of gene transcription in human adipose tissue (66) and GWAS have identified SNPs near/in *KFL14* associated with T2D and HDL cholesterol levels (47,67). Additionally, DNA methylation of a CpG site in the gene body of *SPATA18* correlated positively with age in both adipose tissue from our two cohorts and white blood cells in a previous study by Johansson *et al.* (27). *SPATA18* (*MIEAP*) encodes a p53-inducible protein that controls mitochondrial quality by repairing or eliminating unhealthy mitochondria (36). Maintenance of healthy mitochondria prevents aging, cancer and a variety of degenerative diseases that are due to the result of defective mitochondrial quality control. For *NOX4*, both DNA methylation and mRNA expression were associated with age. Nox4 (NAPDH oxidase 4) has previously been shown to regulate mitochondrial function in an aging-induced senescence model of cultured endothelial cells (68). Interestingly, cellular aging seems to promote Nox4 interaction with mitochondria. This disrupts complex I in the electron transport chain and increases mitochondrial reactive oxygen species (ROS). This could be a contributing factor in the loss of replicative lifespan seen in senescence. A CpG site in the first exon of *PATZ1* correlated with age in adipose tissue of both our cohorts. The POZ/BTB and AT-hook-containing zinc finger protein 1 (PATZ1) seems to have an important role in the regulation of endothelial cell senescence through an ROS-mediated p53-dependent pathway and contribute to vascular diseases associated with aging (35).

It should be noted that our male discovery cohort includes adipose tissue from 96 non-diabetic males with a wide range in all studied phenotypes (age, BMI and HbA1c), whereas our female validation cohort includes 94 non-diabetic females with a smaller age span, but a wide range in BMI and HbA1c. This is a likely explanation for why we identify a smaller number of CpG sites significantly associated with age in the female validation cohort compared with the male discovery cohort. Nevertheless, the majority of CpG sites associated with age in the female validation cohort were also significant in the male discovery cohort, suggesting a general effect of aging on DNA methylation in both genders. Additionally, our male discovery cohort consists of combined data from four different subcohorts (5,9,10,69). Here, we included males without known disease and with DNA available from subcutaneous adipose tissue. It is known that artifacts such as batch effects may reduce the statistical accuracy in genomic data. To adjust for this potential problem, the association between DNA methylation data or mRNA expression data and studied phenotypes was analyzed using a random effect mixed model, including cohort as the random effect variable and age, BMI and HbA1c as fixed factors. We can still not exclude that our data include some false-positive results due to batch effects rather than biological variation. Nevertheless, the fact that we are able to validate many of our results in adipose tissue from a validation cohort as well as in blood from previously published studies strengthens our data. Of note, 9897 methylation sites were significantly associated with age in the same direction both in adipose tissue in our male discovery cohort and in blood in the study by Johansson *et al.* (27).

We also found a strong effect of increased BMI on the degree of DNA methylation in human adipose tissue, indeed proposing that obesity can mediate some of its effects via altering the epigenome. Importantly, DNA methylation of ~5000 CpG sites was associated with BMI in adipose tissue from both our male discovery cohort and female validation cohort. Interestingly, a large number of these CpG sites did also show differential DNA methylation in adipose tissue from subjects with T2D compared with non-diabetic controls (5), suggesting that BMI associated changes in DNA methylation may predispose to T2D. We could also link BMI associated DNA methylation to differential expression of 2825 genes. The strongest correlation between mRNA expression and BMI was seen for *ITIH5*. This gene encodes inter-alpha-trypsin inhibitor heavy chain family member 5 and it is highly expressed in subcutaneous adipose tissue, increased in obesity, down-regulated after weight loss and associated with measures of body size and metabolism (70). Also, eight CpG sites annotated to *ITIH5* correlated with BMI, suggesting a key epigenetic

mechanism for regulation of this gene. Other genes showing both altered DNA methylation and expression in human adipose tissue based on increased BMI include *FTO*, *CCL18*, *MTCH2*, *IRS1* and *SPP1 (OPN)*. Interestingly, we recently found that *CCL18*, encoding CC chemokine ligand 18, and *SPP1*, encoding osteopontin, were the most up-regulated genes in adipose tissue from subjects with T2D compared with non-diabetic controls and both have previously been linked to inflammation (71,72). Additionally, genetic variation in *FTO* has previously been linked to both obesity and T2D, although recent data suggest that *IRX3* rather than *FTO* is mediating the effects of this SNP (73,74). We also identified methylation sites only associated with BMI in the male discovery cohort or the female validation cohort. These differences may be due to gender-specific effects on DNA methylation. Indeed, gender differences in the DNA methylation pattern have previously been reported in human pancreatic islets, the liver, heart muscle, blood and saliva (75–81).

Moreover, a recent investigation related BMI to DNA methylation in whole blood cells from 479 individuals and identified three CpG sites annotated to *HIF3A* with increased DNA methylation associated with increased BMI. The association between BMI and methylation of *HIF3A* was further validated in adipose tissue from 635 females (21). Interestingly, we found that DNA methylation of the same three CpG sites was positively associated with BMI in our female validation cohort, consisting of 94 females, but not in our male discovery cohort, consisting of 96 males. Nevertheless, DNA methylation in two other sites in *HIF3A* was significantly associated with BMI in the 96 males. The protein encoded by *HIF3A*, hypoxia inducible factor 3 alpha subunit, has been shown to play a role in the cellular response to glucose and insulin and to function as an accelerator of adipocyte differentiation (82,83). However, the possible gender-specific relation between *HIF3A* methylation and BMI is novel and has to be validated further.

We also studied the association between HbA1c and DNA methylation in human adipose tissue from non-diabetic subjects. However, the effect of glucose on both DNA methylation and gene expression in human adipose tissue seems less strong than those of BMI and age. Whether this is also the case in other tissues with a key role in whole body glucose homeostasis, i.e. the liver, skeletal muscle and pancreatic islets, remains to be tested. This result may be due to a smaller span in HbA1c compared with age and BMI, since only subjects without known diabetes were included in this study. Importantly, the strong effects of age and BMI on DNA methylation presented in this study can introduce a bias in studies of the DNA methylation pattern in disease, when cases are not carefully matched for age and BMI or when the statistical analyses are not adjusted for these phenotypes, a factor that needs to be considered in the design of future epigenetic studies.

It should also be noted that the impact of age, BMI and HbA1c on the degree of DNA methylation in human adipose tissue in the present study is quite large compared with some previous studies in human adipose tissue where the greatest absolute differences in methylation were ~20% (5,19). Here, we observed larger absolute differences in methylation, e.g. ~40% absolute difference in the methylation of *ELOVL2* between young and elderly subjects and ~25% absolute difference in the methylation of *CCRL2* between lean and obese subjects.

Interestingly, mRNA expression correlated significantly with the degree of DNA methylation for a large proportion of the CpG sites identified in this study. Since all included CpG sites already were shown to be associated with the phenotypes investigated, these results truly supports an interaction between adipose tissue DNA methylation and mRNA expression in establishing metabolic phenotypes.

In addition to adipocytes, adipose tissue comprises a mixture of different cell types, and changes in cell type composition could potentially be responsible for some of the observed changes in DNA methylation. However, as exemplified by the overlap between adipose tissue and blood, some DNA methylation patterns may also be tissue and cell type unspecific. Additionally, when we investigated mRNA expression for cell type-specific markers, no significant associations were found between BMI or HbA1c and *PNPLA2*, *FAS*, *LIPE* and *RETN* as markers of adipocytes, *DLK1* as a marker of preadipocytes, *PRDM16* and *UCP1* as markers of brown adipocytes, *EMR1* as a marker of macrophages, *TNF* and *IL6* representing cytokines and finally *CASP3*, *CASP7* and *LGALS3* as markers for inflammation. We further compared our findings of altered DNA methylation in adipose tissue with cell type-specific methylation sites of candidate genes in inflammatory complex diseases observed in white blood cells (84). Among 8252 analyzed CpG sites in 343 genes, they found 1865 CpG sites differentially methylated between the different cell types in blood. Among our 33 058 CpG sites significantly associated with BMI in adipose tissue, we found 173 CpG sites overlapping with the 1865 cell type-specific methylation sites observed in white blood cells by Reinius *et al.* For HbA1c, only 9 cell type-specific methylation sites were found among our 711 CpG sites associated with HbA1c in adipose tissue. Taken together, these results suggest that there is no major impact of cellular composition or inflammatory response on the observed associations in adipose tissue DNA methylation and BMI or HbA1c. Anyway, although future studies should aim to investigate the DNA methylome in adipocytes isolated from both subcutaneous and intra-abdominal fat tissue, it should be noted that cell isolation processes may alter both gene expression and DNA methylation.

In conclusion, we demonstrate for the first time an impact of age, BMI and HbA1c on the genome-wide DNA methylation pattern in human adipose tissue. Our data support an important function of altered DNA methylation in the development of several non-communicable diseases such as T2D, obesity, cardiovascular disease and cancer. Finally, we demonstrate that epigenetic variation in blood cells can mirror age-related epigenetic signatures in target tissues of important biological function, i.e. adipose tissue. This opens up for the future development and use of blood-based epigenetic biomarkers to predict disease and altered metabolic function.

Materials and Methods

Study participants

The male discovery cohort consists of 96 males from Sweden and Denmark without known disease and with a broad range in age (23–80 years), BMI (17.5–39.0 kg/m^2) and HbA1c levels (28–46 mmol/mol). It should be noted that the range in HbA1c is represented by variation in non-diabetic subjects. This cohort includes males without known disease from four subcohorts, all previously described (5,9,10,69), with DNA available from subcutaneous adipose tissue biopsies taken in the fasted state (19,60,85). Their clinical characteristics are presented in Table 1, and the characteristics of the subjects included in the present study from the four subcohorts are presented separately in Supplementary Material, Table S27. The female validation cohort consists of 94 Swedish females with a broad range in BMI (18.2–44.9 kg/m^2) and HbA1c levels (25–39 mmol/mol), but a more even distribution in age (21–37 years) and with DNA available

from subcutaneous adipose tissue biopsies taken at the fasted state. This cohort is part of a clinical study examining the impact of polycystic ovary syndrome on female metabolism as described previously (32,33). The characteristics of the female validation cohort are shown in Table 1. The mixed validation cohort consists of 37 males and 67 females from Denmark without known disease and with DNA available from subcutaneous adipose tissue biopsies and blood samples taken at the fasted state. This cohort is part of a Danish family study previously described (86,87) and includes 42 families with genetic risk for T2D and with a broad range in age (32–83 years), BMI (18–47 kg/m^2) and HbA1c levels (22–44 mmol/mol). Their characteristics are shown in Table 1.

In all study cohorts, height and weight were measured wearing light clothing and no shoes, and BMI was calculated as weight divided by the square of the height (kg/m^2). Written informed consent was obtained from all participants and the research protocol was approved by the local human research ethics committees.

DNA methylation analysis

DNA was extracted from subcutaneous adipose tissue biopsies and blood using Qiagen DNA extraction kits (Qiagen, Hilden, Germany). DNA methylation was analyzed genome-wide using the Infinium HumanMethylation450 BeadChip assay (Illumina, San Diego, CA, USA), covering a total of 485 577 probes corresponding to 21 231 (99%) RefSeq genes [88,89]. Five hundred nanograms of genomic DNA from adipose tissue was bisulfite-converted using the EZ DNA methylation kit (Zymo Research, Orange, CA, USA) and used with the Infinium® assay, with all other procedures following the standard Infinium HD Assay Methylation Protocol Guide (Part # 15 019 519, Illumina). The Illumina iScan system was used for imaging data on the BeadChips.

GenomeStudio® Methylation module software was used to calculate the raw methylation score for each probe, represented as methylation β-values [β = intensity of the Methylated allele (M)/intensity of the Unmethylated allele (U) + intensity of the Methylated allele (M) + 100]. All samples passed the GenomeStudio quality control steps based on built in control probes for staining, hybridization, extension and specificity, and the bisulfite conversion efficiency was high (intensity signal >4000) (90). The DNA methylation data were exported from GenomeStudio and subsequently analyzed using Bioconductor (91). β-Values were converted to M-values [$M = \log_2(\beta/(1-\beta))$] using the lumi package (92), to make data more homoscedastic and appropriate for further bioinformatical and statistical analyses (93). Next, the probes on the array targeting SNPs (rs; $n = 65$) and non-CpG sites (ch; $n = 2757$) were removed. Additionally, 5448 probes with SNPs in the target CpG (MAF > 0.1 based on dbSNP) (94) and 14 316 probes reported to be cross-reactive with 50 bp (95) were removed. Finally, 6191 probes were filtered away based on Illumina detection P-value (mean $P \geq 0.01$), resulting in a total of 456 800 individual CpG sites from adipose tissue of 96 men for subsequent analyses. After the same quality control and filtering of probes, 460 973 individual CpG sites generated successful DNA methylation data in adipose tissue of 94 women. Of note, probes targeting the Y-chromosome were also filtered away in these women. The DNA methylation data were background corrected by subtracting the median M-value of the 600 built in negative controls and was further normalized using quantile normalization. BMIQ was used to correct for the bias of the two different probe types on the array (96). As 12 samples are analyzed on each Infinium HumanMethylation450 BeadChip, the generated DNA methylation data needed to be batch-corrected. Correction for batch effects within the methylation array data of each cohort was performed using COMBAT (97).

DNA methylation of specific CpG sites was analyzed in bisulfite-treated genomic DNA from adipose tissue and blood cells using pyrosequencing together with the PyroMark PCR kit, PyroMark Gold Q96 reagents and the PyroMark ID 96 (Qiagen) according to the manufacturer's instructions. Primers were designed using the PyroMark Assay design Software 2.0 and data were analyzed with the PyroMark Q96 2.5.7 software program. The following primer sequences were used for pyosequencing of *ELOVL2*: F-primer 5′-GAGGGGAGTAGGGTAAGTGAG-3′, R-primer 5′Biotin-CATTTCCCCCTAATATATACTTCAA-3′, sequencing primer 5′-GGGAGGAGATTTGTAGGTTT-3′, *KLF14*: F-primer 5′-GGTTTTT AGGTTAAGTTATGTTTAATAGT-3′, R-primer 5′Biotin-AAACTA CTACAACCCAAAAATTCC-3′, sequencing primer 5′-ATAGTTTTA GAAATTATTTTGTTT-3′, *CCRL2*: F-primer 5′-AGTTTTAGTTT GGGGTTAAATTTGT-3′, R-primer 5′Biotin-ACAACCAAAAATAA TTAATACTATAACTCA-3′, sequencing primer 5′-ATATTTTTTTTT ATTTAATTTGATG-3′.

mRNA expression analysis

RNA was extracted from subcutaneous adipose tissue biopsies using miRNeasy kit followed by RNeasy MinElute Cleanup kit (Qiagen; subcohort 1, 2 and 3) or using the RNeasy Lipid Tissue Mini Kit (Qiagen; subcohort 4). The BioAnalyzer (Agilent, Santa Clara, CA, USA) was used to measure RNA quality, requiring a RNA integrity number >7 for each sample to be included. Total RNA (200 ng) from subcutaneous adipose tissue biopsies was used for analysis using the GeneChip Human Gene 1.0 ST whole transcript based array (Affymetrix, Santa Clara, CA, USA), following the Affymetrix standard protocol. The Expression Console Software was used for basic Affymetrix chip and experimental quality, and for background correction, data normalization and probe summarization the robust multi-array average method was used (98). Also here, we applied COMBAT to correct for batch effects within each cohort (97).

Statistical analysis

The association between DNA methylation data from the Infinium HumanMethylation450 BeadChip and studied phenotypes was in the male discovery cohort analyzed using a random effect mixed model, including cohort as the random effect variable and age, BMI and HbA1c as fixed factors. Also the association between mRNA expression and age, BMI and HbA1c was analyzed using the same model. In the female validation cohort, the association between DNA methylation data and studied phenotypes was analyzed using the R package 'limma' and linear model including age, BMI, HbA1c and polycystic ovary syndrome status as variables. To account for multiple testing, we applied false discovery rate (FDR) analysis and $q < 0.05$ (FDR < 5%) was considered significant (99). For the overlap between mRNA expression and DNA methylation as well as for pathway analyses, a CpG site was annotated to a gene if it was located between a distance of 500 kb upstream and 100 kb downstream of the gene. Associations between age or BMI and DNA methylation in the mixed validation cohort was analyzed using random effect mixed linear models in R i386 3.1.0 (http://www.r-project.org) and sex, age, BMI and HbA1c were included as fixed factors in all models, whereas family number/pedigree was included as a random factor. Also, associations between DNA methylation levels in adipose tissue and blood in the mixed validation cohort were analyzed with random

effect mixed linear models adjusted for sex (fixed factor) and family number/pedigree (random factor).

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

This study would not be possible without the help from participating clinicians and laboratory technicians, who performed clinical studies and provided the clinical materials, including Targ Elgzyri and Ylva Wessman at Scania University Hospital, Malmö, Sweden, and Marianne Modest and Lars Sander Koch at Steno Diabetes Center, Denmark. We thank SCIBLU (Swegene Center for Integrative Biology at Lund University) Genomics Facility for help with DNA methylation and mRNA expression analyses.

Conflict of Interest statement. None declared.

Funding

This work was supported by grants from the Swedish Research Council, Region Skåne (ALF), Knut and Alice Wallenberg Foundation, Novo Nordisk Foundation, EFSD/Lilly Fellowship, Söderberg Foundation, The Swedish Diabetes foundation, Påhlsson Foundation, EXODIAB, Linné grant (B31 5631/2006), The Danish Strategic Research Council, The Danish Council for Independent Research, Rigshospitalet, University of Copenhagen, Steno Diabetes Center, Danish Diabetes Academy, Jane and Dan Olsson Foundation, Swedish federal government under the LUA/ALF agreement ALFGBG-136481 and the Regional Research and Development agreement (VGFOUREG-5171, –11296, and –7861). The study behind the mixed validation cohort was supported by the Lundbeck Foundation (The Lundbeck Foundation Centre for Applied Medical Genomics in Personalised Disease Prediction, Prevention and Care [LuCamp], www.lucamp.org) and The Danish Council for Independent Research. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk).

References

1. Ling, C. and Groop, L. (2009) Epigenetics: a molecular link between environmental factors and type 2 diabetes. *Diabetes*, **58**, 2718–2725.
2. Rakyan, V.K., Down, T.A., Balding, D.J. and Beck, S. (2011) Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.*, **12**, 529–541.
3. Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suner, D., Cigudosa, J.C., Urioste, M., Benitez, J. *et al.* (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA*, **102**, 10604–10609.
4. Kaminsky, Z.A., Tang, T., Wang, S.C., Ptak, C., Oh, G.H., Wong, A.H., Feldcamp, L.A., Virtanen, C., Halfvarson, J., Tysk, C. *et al.* (2009) DNA methylation profiles in monozygotic and dizygotic twins. *Nat. Genet.*, **41**, 240–245.
5. Nilsson, E., Jansson, P.A., Perfilyev, A., Volkov, P., Pedersen, M., Svensson, M.K., Poulsen, P., Ribel-Madsen, R., Pedersen, N.L., Almgren, P. *et al.* (2014) Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes*, **63**, 2962–2976.
6. Ribel-Madsen, R., Fraga, M.F., Jacobsen, S., Bork-Jensen, J., Lara, E., Calvanese, V., Fernandez, A.F., Friedrichsen, M., Vind, B.F., Hojlund, K. *et al.* (2012) Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. *PLoS One*, **7**, e51302.
7. Chen, L., Magliano, D.J. and Zimmet, P.Z. (2012) The worldwide epidemiology of type 2 diabetes mellitus–present and future perspectives. *Nat. Rev. Endocrinol.*, **8**, 228–236.
8. Barker, D.J. (1997) Maternal nutrition, fetal nutrition, and disease in later life. *Nutrition*, **13**, 807–813.
9. Brons, C., Jensen, C.B., Storgaard, H., Alibegovic, A., Jacobsen, S., Nilsson, E., Astrup, A., Quistorff, B. and Vaag, A. (2008) Mitochondrial function in skeletal muscle is normal and unrelated to insulin action in young men born with low birth weight. *J. Clin. Endocrinol. Metab.*, **93**, 3885–3892.
10. Jørgensen, S.W., Brøns, C., Bluck, L., Hjort, L., Færch, K., Thankamony, A., Gillberg, L., Friedrichsen, M., Dunger, D.B. and Vaag, A. (2015) Metabolic response to 36 hours of fasting in young men born small vs appropriate for gestational age. *Diabetologia*, **158**, 178–187.
11. Garg, N., Moorthy, N., Kapoor, A., Tewari, S., Kumar, S., Sinha, A., Shrivastava, A. and Goel, P.K. (2014) Hemoglobin A(1c) in nondiabetic patients: an independent predictor of coronary artery disease and its severity. *Mayo Clin. Proc.*, **89**, 908–916.
12. Global Burden of Metabolic Risk Factors for Chronic Diseases, C.Lu, Y., Hajifathalian, K., Ezzati, M., Woodward, M., Rimm, E.B. and Danaei, G. (2014) Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. *Lancet*, **383**, 970–983.
13. Lyssenko, V., Jonsson, A., Almgren, P., Pulizzi, N., Isomaa, B., Tuomi, T., Berglund, G., Altshuler, D., Nilsson, P. and Groop, L. (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N. Engl. J. Med.*, **359**, 2220–2232.
14. de Beer, J.C. and Liebenberg, L. (2014) Does cancer risk increase with HbA1c, independent of diabetes? *Br. J. Cancer*, **110**, 2361–2368.
15. Liu, L., van Groen, T., Kadish, I., Li, Y., Wang, D., James, S.R., Karpf, A.R. and Tollefsbol, T.O. (2011) Insufficient DNA methylation affects healthy aging and promotes age-related health problems. *Clin. Epigenetics*, **2**, 349–360.
16. Selvin, E., Steffes, M.W., Zhu, H., Matsushita, K., Wagenknecht, L., Pankow, J., Coresh, J. and Brancati, F.L. (2010) Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *N. Engl. J. Med.*, **362**, 800–811.
17. Ronti, T., Lupattelli, G. and Mannarino, E. (2006) The endocrine function of adipose tissue: an update. *Clin. Endocrinol. (Oxf)*, **64**, 355–365.
18. Bays, H.E. (2012) Adiposopathy, diabetes mellitus, and primary prevention of atherosclerotic coronary artery disease: treating "sick fat" through improving fat function with antidiabetes therapies. *Am. J. Cardiol.*, **110**, 4B–12B.
19. Rönn, T., Volkov, P., Davegardh, C., Dayeh, T., Hall, E., Olsson, A.H., Nilsson, E., Tornberg, A., Dekker Nitert, M., Eriksson, K.F. *et al.* (2013) A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. *PLoS Genet.*, **9**, e1003572.
20. Rönn, T., Volkov, P., Tornberg, A., Elgzyri, T., Hansson, O., Eriksson, K.F., Groop, L. and Ling, C. (2014) Extensive changes in the transcriptional profile of human adipose tissue including genes involved in oxidative phosphorylation after a 6-month exercise intervention. *Acta Physiol. (Oxf)*, **211**, 188–200.

21. Dick, K.J., Nelson, C.P., Tsaprouni, L., Sandling, J.K., Aissi, D., Wahl, S., Meduri, E., Morange, P.E., Gagnon, F., Grallert, H. *et al.* (2014) DNA methylation and body-mass index: a genome-wide analysis. *Lancet*, **383**, 1990–1998.
22. Teschendorff, A.E., West, J. and Beck, S. (2013) Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum. Mol. Genet.*, **22**, R7–R15.
23. Dayeh, T., Volkov, P., Salo, S., Hall, E., Nilsson, E., Olsson, A.H., Kirkpatrick, C.L., Wollheim, C.B., Eliasson, L., Ronn, T. *et al.* (2014) Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet.*, **10**, e1004160.
24. Ling, C., Poulsen, P., Simonsson, S., Rönn, T., Holmkvist, J., Almgren, P., Hagert, P., Nilsson, E., Mabey, A.G., Nilsson, P. *et al.* (2007) Genetic and epigenetic factors are associated with expression of respiratory chain component NDUFB6 in human skeletal muscle. *J. Clin. Invest.*, **117**, 3427–3435.
25. Rönn, T., Poulsen, P., Hansson, O., Holmkvist, J., Almgren, P., Nilsson, P., Tuomi, T., Isomaa, B., Groop, L., Vaag, A. *et al.* (2008) Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle. *Diabetologia*, **51**, 1159–1168.
26. Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y. *et al.* (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell*, **49**, 359–367.
27. Johansson, A., Enroth, S. and Gyllensten, U. (2013) Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS One*, **8**, e67378.
28. Florath, I., Butterbach, K., Muller, H., Bewerunge-Hudler, M. and Brenner, H. (2013) Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over novel age-associated CpG sites. *Hum. Mol. Genet.*, **23**, 1186–1201.
29. Day, K., Waite, L.L., Thalacker-Mercer, A., West, A., Bamman, M.M., Brooks, J.D., Myers, R.M. and Absher, D. (2013) Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biol.*, **14**, R102.
30. Fox, J. and Monette, G. (1992) Generalized collinearity diagnostics. *J. Am. Statist. Assoc.*, **87**, 178–183.
31. Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, **13**, 484–492.
32. Manneras-Holm, L., Leonhardt, H., Kullberg, J., Jennische, E., Oden, A., Holm, G., Hellstrom, M., Lonn, L., Olivecrona, G., Stener-Victorin, E. *et al.* (2011) Adipose tissue has aberrant morphology and function in PCOS: enlarged adipocytes and low serum adiponectin, but not circulating sex steroids, are strongly associated with insulin resistance. *J. Clin. Endocrinol. Metab.*, **96**, E304–E311.
33. Stener-Victorin, E., Holm, G., Labrie, F., Nilsson, L., Janson, P.O. and Ohlsson, C. (2010) Are there any sensitive and specific sex steroid markers for polycystic ovary syndrome? *J. Clin. Endocrinol. Metab.*, **95**, 810–819.
34. Calabrese, F., Guidotti, G., Racagni, G. and Riva, M.A. (2013) Reduced neuroplasticity in aged rats: a role for the neurotrophin brain-derived neurotrophic factor. *Neurobiol. Aging*, **34**, 2768–2776.
35. Cho, J.H., Kim, M.J., Kim, K.J. and Kim, J.R. (2012) POZ/BTB and AT-hook-containing zinc finger protein 1 (PATZ1) inhibits endothelial cell senescence through a p53 dependent pathway. *Cell Death Differ.*, **19**, 703–712.
36. Kitamura, N., Nakamura, Y., Miyamoto, Y., Miyamoto, T., Kabu, K., Yoshida, M., Futamura, M., Ichinose, S. and Arakawa, H. (2011) Mieap, a p53-inducible protein, controls mitochondrial quality by repairing or eliminating unhealthy mitochondria. *PLoS One*, **6**, e16060.
37. Makpol, S., Zainuddin, A., Chua, K.H., Mohd Yusof, Y.A. and Ngah, W.Z. (2013) Gamma-tocotrienol modulated gene expression in senescent human diploid fibroblasts as revealed by microarray analysis. *Oxid. Med. Cell Longev.*, **2013**, 454328.
38. Nakura, J., Miki, T., Nagano, K., Kihara, K., Ye, L., Kamino, K., Fujiwara, Y., Yoshida, S., Murano, S., Fukuchi, K. *et al.* (1993) Close linkage of the gene for Werner's syndrome to ANK1 and D8S87 on the short arm of chromosome 8. *Gerontology*, **39**(Suppl 1), 11–15.
39. Steegenga, W.T., Boekschoten, M.V., Lute, C., Hooiveld, G.J., de Groot, P.J., Morris, T.J., Teschendorff, A.E., Butcher, L.M., Beck, S. and Muller, M. (2014) Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr)*, **36**, 9648.
40. Bell, J.T., Tsai, P.C., Yang, T.P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A. *et al.* (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.*, **8**, e1002629.
41. Garagnani, P., Bacalini, M.G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., Di Blasio, A.M., Gentilini, D., Vitale, G., Collino, S. *et al.* (2012) Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*, **11**, 1132–1134.
42. Heyn, H., Li, N., Ferreira, H.J., Moran, S., Pisano, D.G., Gomez, A., Diez, J., Sanchez-Mut, J.V., Setien, F., Carmona, F.J. *et al.* (2012) Distinct DNA methylomes of newborns and centenarians. *Proc. Natl Acad. Sci. USA*, **109**, 10522–10527.
43. Rakyan, V.K., Down, T.A., Maslau, S., Andrew, T., Yang, T.P., Beyan, H., Whittaker, P., McCann, O.T., Finer, S., Valdes, A.M. *et al.* (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.*, **20**, 434–439.
44. Teschendorff, A.E., Menon, U., Gentry-Maharaj, A., Ramus, S.J., Weisenberger, D.J., Shen, H., Campan, M., Noushmehr, H., Bell, C.G., Maxwell, A.P. *et al.* (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.*, **20**, 440–446.
45. Xu, Z. and Taylor, J.A. (2014) Genome-wide age-related DNA methylation changes in blood and other tissues relate to histone modification, expression and cancer. *Carcinogenesis*, **35**, 356–364.
46. Laakso, M., Malkki, M., Kekalainen, P., Kuusisto, J. and Deeb, S.S. (1994) Insulin receptor substrate-1 variants in non-insulin-dependent diabetes. *J. Clin. Invest.*, **94**, 1141–1146.
47. Wang, J., Zhang, J., Shen, J., Hu, D., Yan, G., Liu, X., Xu, X., Pei, L., Li, Y. and Sun, C. (2014) Association of KCNQ1 and KLF14 polymorphisms and risk of type 2 diabetes mellitus: a global meta-analysis. *Hum. Immunol.*, **75**, 342–347.
48. Arner, P., Bernard, S., Salehpour, M., Possnert, G., Liebl, J., Steier, P., Buchholz, B.A., Eriksson, M., Arner, E., Hauner, H. *et al.* (2011) Dynamics of human adipose lipid turnover in health and metabolic disease. *Nature*, **478**, 110–113.
49. El-Osta, A., Brasacchio, D., Yao, D., Pocai, A., Jones, P.L., Roeder, R.G., Cooper, M.E. and Brownlee, M. (2008) Transient high glucose causes persistent epigenetic changes and altered gene expression during subsequent normoglycemia. *J. Exp. Med.*, **205**, 2409–2417.
50. Reddy, M.A. and Natarajan, R. (2013) Role of epigenetic mechanisms in the vascular complications of diabetes. *Subcell. Biochem.*, **61**, 435–454.

51. Yang, B.T., Dayeh, T.A., Kirkpatrick, C.L., Taneera, J., Kumar, R., Groop, L., Wollheim, C.B., Nitert, M.D. and Ling, C. (2011) Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. *Diabetologia*, **54**, 360–367.
52. Yang, B.T., Dayeh, T.A., Volkov, P.A., Kirkpatrick, C.L., Malmgren, S., Jing, X., Renstrom, E., Wollheim, C.B., Nitert, M.D. and Ling, C. (2012) Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. *Mol. Endocrinol.*, **26**, 1203–1202.
53. Dasu, M.R., Devaraj, S., Park, S. and Jialal, I. (2010) Increased toll-like receptor (TLR) activation and TLR ligands in recently diagnosed type 2 diabetic subjects. *Diabetes Care*, **33**, 861–868.
54. Ljubicic, S., Bezzi, P., Brajkovic, S., Nesca, V., Guay, C., Ohbayashi, N., Fukuda, M., Abderrhamani, A. and Regazzi, R. (2013) The GTPase Rab37 participates in the control of insulin exocytosis. *PLoS One*, **8**, e68255.
55. Varney, M.D., Valdes, A.M., Carlson, J.A., Noble, J.A., Tait, B.D., Bonella, P., Lavant, E., Fear, A.L., Louey, A., Moonsamy, P. *et al.* (2010) HLA DPA1, DPB1 alleles and haplotypes contribute to the risk associated with type 1 diabetes: analysis of the type 1 diabetes genetics consortium families. *Diabetes*, **59**, 2055–2062.
56. Rakyan, V.K., Down, T.A., Thorne, N.P., Flicek, P., Kulesha, E., Graf, S., Tomazou, E.M., Backdahl, L., Johnson, N., Herberth, M. *et al.* (2008) An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.*, **18**, 1518–1529.
57. Geeleher, P., Hartnett, L., Egan, L.J., Golden, A., Raja Ali, R.A. and Seoighe, C. (2013) Gene-set analysis is severely biased when applied to genome-wide methylation data. *Bioinformatics*, **29**, 1851–1857.
58. Chen, H., Kazemier, H.G., de Groote, M.L., Ruiters, M.H., Xu, G.L. and Rots, M.G. (2013) Induced DNA demethylation by targeting ten-eleven translocation 2 to the human ICAM-1 promoter. *Nucleic Acids Res.*, **8**, 203–209.
59. Brons, C., Jacobsen, S., Nilsson, E., Ronn, T., Jensen, C.B., Storgaard, H., Poulsen, P., Groop, L., Ling, C., Astrup, A. *et al.* (2010) Deoxyribonucleic acid methylation and gene expression of PPARGC1A in human muscle is influenced by high-fat overfeeding in a birth-weight-dependent manner. *J. Clin. Endocrinol. Metab.*, **95**, 3048–3056.
60. Gillberg, L., Jacobsen, S.C., Rönn, T., Brons, C. and Vaag, A. (2014) PPARGC1A DNA methylation in subcutaneous adipose tissue in low birth weight subjects—impact of 5 days of high-fat overfeeding. *Metabolism*, **63**, 263–271.
61. Hall, E., Volkov, P., Dayeh, T., Bacos, K., Ronn, T., Nitert, M.D. and Ling, C. (2014) Effects of palmitate on genome-wide mRNA expression and DNA methylation patterns in human pancreatic islets. *BMC Med.*, **12**, 103.
62. Nitert, M.D., Dayeh, T., Volkov, P., Elgzyri, T., Hall, E., Nilsson, E., Yang, B.T., Lang, S., Parikh, H., Wessman, Y. *et al.* (2012) Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. *Diabetes*, **61**, 3322–3332.
63. Nilsson, E.E. and Skinner, M.K. (2015) Environmentally induced epigenetic transgenerational inheritance of disease susceptibility. *Transl. Res.*, **165**, 12–17.
64. Simmons, R. (2011) Epigenetics and maternal nutrition: nature v. nurture. *Proc. Nutr. Soc.*, **70**, 73–81.
65. Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E. *et al.* (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature*, **500**, 477–481.
66. Small, K.S., Hedman, A.K., Grundberg, E., Nica, A.C., Thorleifsson, G., Kong, A., Thorsteindottir, U., Shin, S.Y., Richards, H.B., Soranzo, N. *et al.* (2011) Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat. Genet.*, **43**, 561–564.
67. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J. *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707–713.
68. Koziel, R., Pircher, H., Kratochwil, M., Lener, B., Hermann, M., Dencher, N.A. and Jansen-Durr, P. (2013) Mitochondrial respiratory chain complex I is inactivated by NADPH oxidase Nox4. *Biochem. J.*, **452**, 231–239.
69. Elgzyri, T., Parikh, H., Zhou, Y., Nitert, M.D., Ronn, T., Segerstrom, A.B., Ling, C., Franks, P.W., Wollmer, P., Eriksson, K.F. *et al.* (2012) First-degree relatives of type 2 diabetic patients have reduced expression of genes involved in fatty acid metabolism in skeletal muscle. *J. Clin. Endocrinol. Metab.*, **97**, 1332–1337.
70. Anveden, A., Sjoholm, K., Jacobson, P., Palsdottir, V., Walley, A.J., Froguel, P., Al-Daghri, N., McTernan, P.G., Mejhert, N., Arner, P. *et al.* (2012) ITIH-5 expression in human adipose tissue is increased in obesity. *Obesity (Silver Spring)*, **20**, 708–714.
71. Mazzali, M., Kipari, T., Ophascharoensuk, V., Wesson, J.A., Johnson, R. and Hughes, J. (2002) Osteopontin—a molecule for all seasons. *QJM*, **95**, 3–13.
72. Schutyser, E., Richmond, A. and Van Damme, J. (2005) Involvement of CC chemokine ligand 18 (CCL18) in normal and pathological processes. *J. Leukoc. Biol.*, **78**, 14–26.
73. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W. *et al.* (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889–894.
74. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gomez-Marin, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F. *et al.* (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, **507**, 371–375.
75. Boks, M.P., Derks, E.M., Weisenberger, D.J., Strengman, E., Janson, E., Sommer, I.E., Kahn, R.S. and Ophoff, R.A. (2009) The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One*, **4**, e6767.
76. Cotton, A.M., Lam, L., Affleck, J.G., Wilson, I.M., Penaherrera, M.S., McFadden, D.E., Kobor, M.S., Lam, W.L., Robinson, W.P. and Brown, C.J. (2011) Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. *Hum. Genet.*, **130**, 187–201.
77. Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A. *et al.* (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.*, **38**, 1378–1385.
78. El-Maarri, O., Becker, T., Junen, J., Manzoor, S.S., Diaz-Lacava, A., Schwaab, R., Wienker, T. and Oldenburg, J. (2007) Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. *Hum. Genet.*, **122**, 505–514.
79. Hall, E., Volkov, P., Dayeh, T., Esguerra, J.L., Salo, S., Eliasson, L., Ronn, T., Bacos, K. and Ling, C. (2014) Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biol.*, **15**, 522.

80. McCarthy, N.S., Melton, P.E., Cadby, G., Yazar, S., Franchina, M., Moses, E.K., Mackey, D.A. and Hewitt, A.W. (2014) Meta-analysis of human methylation data for evidence of sex-specific autosomal patterns. *BMC Genomics*, **15**, 981.
81. Sarter, B., Long, T.I., Tsong, W.H., Koh, W.P., Yu, M.C. and Laird, P.W. (2005) Sex differential in methylation patterns of selected genes in Singapore Chinese. *Hum. Genet.*, **117**, 402–403.
82. Hatanaka, M., Shimba, S., Sakaue, M., Kondo, Y., Kagechika, H., Kokame, K., Miyata, T. and Hara, S. (2009) Hypoxia-inducible factor-3alpha functions as an accelerator of 3T3-L1 adipose differentiation. *Biol. Pharm. Bull.*, **32**, 1166–1172.
83. Heidbreder, M., Qadri, F., Johren, O., Dendorfer, A., Depping, R., Frohlich, F., Wagner, K.F. and Dominiak, P. (2007) Non-hypoxic induction of HIF-3alpha by 2-deoxy-D-glucose and insulin. *Biochem. Biophys. Res. Commun.*, **352**, 437–443.
84. Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S.E., 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**, e41361.
85. Jacobsen, S.C., Gillberg, L., Bork-Jensen, J., Ribel-Madsen, R., Lara, E., Calvanese, V., Ling, C., Fernandez, A.F., Fraga, M.F., Poulsen, P. *et al.* (2014) Young men with low birthweight exhibit decreased plasticity of genome-wide muscle DNA methylation by high-fat overfeeding. *Diabetologia*, **57**, 1154–1158.
86. Boesgaard, T.W., Gjesing, A.P., Grarup, N., Rutanen, J., Jansson, P.A., Hribal, M.L., Sesti, G., Fritsche, A., Stefan, N., Staiger, H. *et al.* (2009) Variant near ADAMTS9 known to associate with type 2 diabetes is related to insulin resistance in offspring of type 2 diabetes patients—EUGENE2 study. *PLoS One*, **4**, e7236.
87. Gillberg, L., Jacobsen, S., Ribel-Madsen, R., Gjesing, A.P., Boesgaard, T.W., Ling, C., Pedersen, O., Hansen, T. and Vaag, A. (2013) Does DNA methylation of PPARGC1A influence insulin action in first degree relatives of patients with type 2 diabetes? *PLoS One*, **8**, e58384.
88. Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., Fan, J.B. and Shen, R. (2011) High density DNA methylation array with single CpG site resolution. *Genomics*, **98**, 288–295.
89. Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C. and Fuks, F. (2011) Evaluation of the Infinium Methylation 450K technology. *Epigenomics*, **3**, 771–784.
90. Teschendorff, A.E., Menon, U., Gentry-Maharaj, A., Ramus, S.J., Gayther, S.A., Apostolidou, S., Jones, A., Lechner, M., Beck, S., Jacobs, I.J. *et al.* (2009) An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One*, **4**, e8274.
91. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
92. Du, P., Kibbe, W.A. and Lin, S.M. (2008) lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, **24**, 1547–1548.
93. Du, P., Zhang, X., Huang, C.C., Jafari, N., Kibbe, W.A., Hou, L. and Lin, S.M. (2010) Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, **11**, 587.
94. Price, M.E., Cotton, A.M., Lam, L.L., Farre, P., Emberly, E., Brown, C.J., Robinson, W.P. and Kobor, M.S. (2013) Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin*, **6**, 4.
95. Chen, Y.A., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J. and Weksberg, R. (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*, **8**, 203–209.
96. Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D. and Beck, S. (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*, **29**, 189–196.
97. Johnson, W.E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.
98. Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**(2), 249–264.
99. Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.

Study III

PLOS ONE

RESEARCH ARTICLE

A Genome-Wide mQTL Analysis in Human Adipose Tissue Identifies Genetic Variants Associated with DNA Methylation, Gene Expression and Metabolic Traits

Petr Volkov[1], Anders H. Olsson[1,2], Linn Gillberg[2], Sine W. Jørgensen[2], Charlotte Brøns[2], Karl-Fredrik Eriksson[3], Leif Groop[4], Per-Anders Jansson[5], Emma Nilsson[1,2], Tina Rönn[1], Allan Vaag[2], Charlotte Ling[1]*

1 Department of Clinical Sciences, Epigenetics and Diabetes, Lund University Diabetes Centre, Clinical Research Centre, Malmö, Sweden, **2** Department of Endocrinology, Diabetes and Metabolism, Rigshospitalet, Copenhagen, Denmark, **3** Department of Clinical Sciences, Vascular Diseases, Lund University, Malmö, Sweden, **4** Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Clinical Research Centre, Malmö, Sweden, **5** The Lundberg Laboratory for Diabetes Research, Center of Excellence for Cardiovascular and Metabolic Research, Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska University Hospital, University of Gothenburg, Gothenburg, Sweden

* charlotte.ling@med.lu.se

Citation: Volkov P, Olsson AH, Gillberg L, Jørgensen SW, Brøns C, Eriksson K-F, et al. (2016) A Genome-Wide mQTL Analysis in Human Adipose Tissue Identifies Genetic Variants Associated with DNA Methylation, Gene Expression and Metabolic Traits. PLoS ONE 11(6): e0157776. doi:10.1371/journal.pone.0157776

Editor: Tanja Zeller, Medical University Hamburg, University Heart Center, GERMANY

Received: October 23, 2015

Accepted: June 3, 2016

Published: June 20, 2016

Data Availability Statement: All data files are available from the GEO database.

Funding: This work was supported by grants from the Swedish Research Council (2013/3018, http://www.vr.se/, CL; 523-2010-1062, http://www.vr.se/, CL), Region Skåne (ALF), Knut and Alice Wallenberg Foundation, Novo Nordisk Foundation, EFSD/Lilly Fellowship, Söderberg Foundation, The Swedish Diabetes foundation, Påhlsson Foundation, EXODIAB, Linné grant (B31 5631/2006), The Danish Strategic Research Council, The Danish Council for

Abstract

Little is known about the extent to which interactions between genetics and epigenetics may affect the risk of complex metabolic diseases and/or their intermediary phenotypes. We performed a genome-wide DNA methylation quantitative trait locus (mQTL) analysis in human adipose tissue of 119 men, where 592,794 single nucleotide polymorphisms (SNPs) were related to DNA methylation of 477,891 CpG sites, covering 99% of RefSeq genes. SNPs in significant mQTLs were further related to gene expression in adipose tissue and obesity related traits. We found 101,911 SNP-CpG pairs (mQTLs) in *cis* and 5,342 SNP-CpG pairs in *trans* showing significant associations between genotype and DNA methylation in adipose tissue after correction for multiple testing, where *cis* is defined as distance less than 500 kb between a SNP and CpG site. These mQTLs include reported obesity, lipid and type 2 diabetes loci, e.g. *ADCY3*/*POMC*, *APOA5*, *CETP*, *FADS2*, *GCKR*, *SORT1* and *LEPR*. Significant mQTLs were overrepresented in intergenic regions meanwhile underrepresented in promoter regions and CpG islands. We further identified 635 SNPs in significant *cis*-mQTLs associated with expression of 86 genes in adipose tissue including *CHRNA5*, *G6PC2*, *GPX7*, *RPL27A*, *THNSL2* and *ZFP57*. SNPs in significant mQTLs were also associated with body mass index (BMI), lipid traits and glucose and insulin levels in our study cohort and public available consortia data. Importantly, the Causal Inference Test (CIT) demonstrates how genetic variants mediate their effects on metabolic traits (e.g. BMI, cholesterol, high-density lipoprotein (HDL), hemoglobin A1c (HbA1c) and homeostatic model assessment of insulin resistance (HOMA-IR)) via altered DNA methylation in human adipose tissue. This study identifies genome-wide interactions between genetic and epigenetic variation in both

Independent Research, Rigshospitalet, University of Copenhagen, Steno Diabetes Center, and Danish Diabetes Academy. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: mQTL, methylation quantitative trait locus; SNP, single nucleotide polymorphism; BMI, body mass index; CIT, causal inference test; HbA1c, hemoglobin A1c; HDL, high-density lipoprotein; LDL, low density lipoprotein; HOMA-IR, homeostasis model insulin resistance; HOMA-B, homeostasis model beta-cell function; CRP, C-reactive protein; GWAS, genome-wide association study; TSS, transcription start site; UTR, untranslated region; WHR, waist-hip ratio; LD, linkage disequilibrium; KEGG, Kyoto Encyclopaedia of Genes and Genomes.

cis and *trans* positions influencing gene expression in adipose tissue and *in vivo* (dys)metabolic traits associated with the development of obesity and diabetes.

Introduction

Genetic factors contribute to the risk of complex metabolic diseases such as obesity and type 2 diabetes. Although genome-wide association studies (GWAS) have identified numerous genetic loci influencing the risk of developing obesity and type 2 diabetes, only a few of these loci have been linked to the molecular mechanisms contributing to the phenotype outcome [1]. Moreover, the identified genetic loci do only explain a modest proportion of the estimated heritability of these diseases and additional genetic mechanisms remain to be found. These may include genetic variants interacting with epigenetic modifications.

The phenomenon of epigenetic modifications are of interest to study for their possible involvement in phenotype transmission and predisposition to complex human diseases, including obesity and type 2 diabetes [2,3]. Epigenetics has been defined as heritable changes in gene function that occur without alterations in the DNA sequence and includes the molecular mechanism of DNA methylation [4]. In differentiated mammalian cells, DNA methylation occurs primarily at cytosines in CG dinucleotides, so called CpG methylation, which is associated with regulation of cell specific gene expression [5,6]. DNA methylation patterns are mainly established early in life, but may also be dynamic and change in response to environmental stimulations such as diet and exercise [7–10]. Concurrently, once epigenetic modifications are introduced they can be stable and inherited [11,12], making epigenetics a potentially important pathogenic mechanism in complex metabolic diseases. Interestingly, twin studies provide evidence for an underlying genetic effect on DNA methylation patterns [13–16]. For example using monozygotic and dizygotic twins, Grundberg et al showed that as much as 37% of the methylation variance can be attributed to genetic factors, which is in line with previous studies [15,16]. In addition, recent studies showed that common genetic variation regulates DNA methylation levels, so called methylation quantitative trait loci (mQTLs) [16–20]. However, most of these studies have been limited to analyses of ~0.1% of human CpG sites in promoter regions [17–19] or restricted to SNPs located within 100 kb from analyzed CpG sites [16]. It remains to be tested if genetic and epigenetic variation interacts throughout the genome in human adipose tissue and subsequently affect gene expression and metabolic traits such as BMI, lipid levels and hemoglobin A1c (HbA1c) in the studied individuals.

The aim of the present study was therefore to perform a genome-wide mQTL analysis in human adipose tissue, investigating both *cis* and *trans* effects of genetic variation on DNA methylation covering most genes and regions in the human genome. Identified mQTLs were followed-up and related to gene expression in adipose tissue. Additionally, since the adipose tissue contributes to whole body energy homeostasis by glucose uptake, triglyceride storage and adipokine secretion, we investigated if the identified SNPs in significant mQTLs affect metabolic traits that are associated with increased risk of obesity and type 2 diabetes in the studied cohort. We further used a causal inference test (CIT) [21] to model the potential causal relationships between genotype, DNA methylation and metabolic phenotypes.

The present study provides the first detailed map of genetic loci in both *cis* and *trans* positions affecting the genome-wide DNA methylation pattern in human adipose tissue as well as numerous metabolic traits. Identified mQTLs cover known lipid, obesity and diabetes loci. Our study highlights that interaction analysis between genetic and epigenetic variation in a tissue of

relevance for metabolic diseases may give new insights to biological processes affecting disease susceptibility.

Results

Associations between genetic variation and DNA methylation in human adipose tissue–a genome-wide mQTL analysis

To examine and map underlying genetic control of DNA methylation patterns in human adipose tissue, we performed a genome-wide mQTL analysis (**Fig 1**). While most previous mQTL studies have been limited to analysis of ~0.1% of human CpG sites [17–19] or SNPs within 100 kb from analyzed CpG sites [16] we performed the first combined *cis*- and *trans*-mQTL analysis covering DNA methylation of most genes and genomic regions in human adipose tissue of 119 Scandinavian men (**Table 1**). Here, we pairwise associated genotype data of 592,794 common SNPs (MAF>0.05) with DNA methylation of 477,891 CpG sites throughout the human genome using a linear regression model including sub-cohort, age and BMI as covariates.

The *cis*-mQTL analysis was limited to SNPs located within 500 kb of either side of the analyzed CpG sites. Here, we detected 101,911 SNP-CpG pairs (mQTLs) showing significant associations between genotype and the degree of DNA methylation after correction for multiple

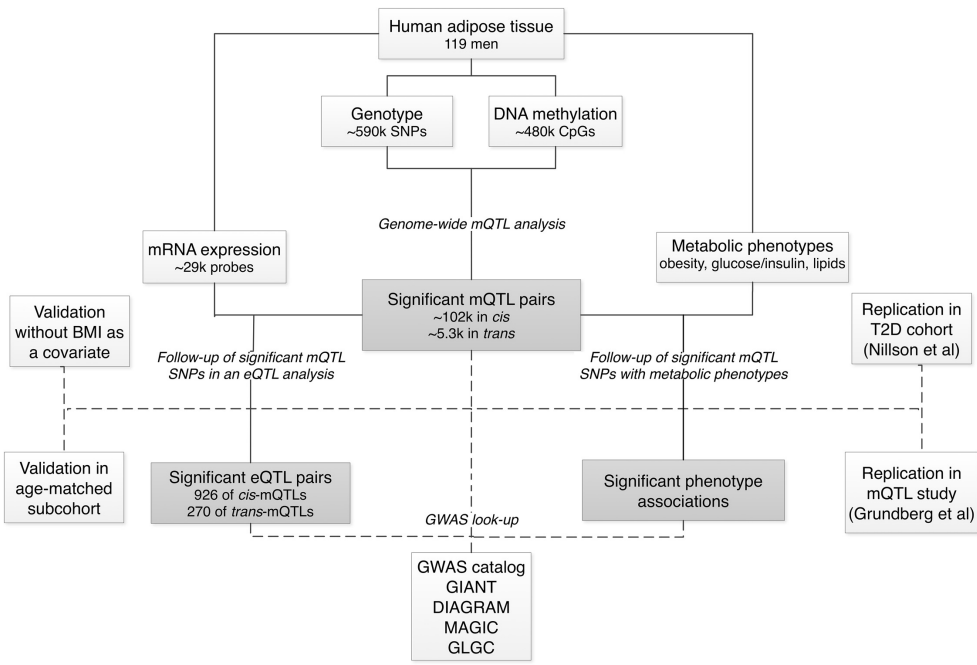


Fig 1. Analysis flowchart of the study.

doi:10.1371/journal.pone.0157776.g001

Table 1. Characteristics of 119 Scandinavian men included in the mQTL analysis.

Phenotype	Mean ± SD	Min	1st. quartile	Median	3rd quartile	Max
Age (years)	31.03 ± 12.3	22	24	25	35	80
Fasting Glucose (mmol/l)	4.76 ± 0.64	3.2	4.4	4.7	5	7
Fasting Insulin (pmol/l)	37.29 ± 22.43	8	23.1	33	43.3	181.3
Weight (kg)	80.86 ± 11.6	57.2	72.6	80.4	89.57	112.7
BMI (kg/m^2)	24.91 ± 3.7	16.4	22.2	24.6	27.15	39
Waist (cm)	90.31 ± 11.5	68	80.75	91	98.25	129
Hip (cm)	97.55 ± 8.8	78	91.75	98	104.2	113
Waist-Hip ratio	0.9 ± 0.06	0.79	0.87	0.9	0.92	1
Cholesterol (mmol/l)	4.5 ± 0.84	2.1	3.9	4.5	5.1	7.1
Triglycerides (mmol/l)	1.14 ± 0.66	0.3	0.72	1	1.3	4.9
HDL (mmol/l)	1.16 ± 0.2	0.5	1	1.13	1.37	1.86
LDL (mmol/l)	2.8 ± 0.77	1	2.3	2.8	3.5	4.7
HbA1c (%)	4.93 ± 0.48	3.7	4.7	5	5.2	6.4
HOMA-IR	1.15 ± 0.78	0.2	0.7	1	1.4	6.5
HOMA-B	133.69 ± 226.33	19.2	56	75.6	118.9	1834

doi:10.1371/journal.pone.0157776.t001

testing (**see** Methods), corresponding to 51,143 unique SNPs and 15,208 unique CpG sites (**Table 2** and **S1 Table**). Of these 15,208 significant CpG sites, 10,064 were annotated to 5,589 unique genes (**Table 2**) and 5,144 CpG sites were annotated to intergenic regions. The most and least significant *cis*-mQTLs are shown in **Fig 2A–2B.**

Previously, we reported that approximately 50% of type 2 diabetes associated SNPs identified by GWAS either introduce or remove a CpG site, a so called CpG-SNP. These CpG-SNPs were further associated with differential DNA methylation of the CpG-SNP site in human pancreatic islets [22]. Among the significant *cis*-mQTLs in the present study, 447 SNPs were located within a CpG site, i.e. the distance between a SNP and CpG site is 0 or 1 and thereby remove or introduce a CpG site–CpG-SNPs (**S1 Table**). The most significant mQTL among these 447 *cis*-mQTLs is presented in **Fig 2C.**

When a distance analysis was performed, we found an overrepresentation ($p < 2.2^{-16}$) of SNPs in significant *cis*-mQTL located close to the CpG site (**Fig 2D–2E**), with a median distance between SNPs and CpG sites of significant *cis*-mQTLs of 29.6 kb. Moreover, the strongest association signals were found for SNPs located close to a CpG site (**Fig 2F**).

In the *trans*-mQTL analysis, including SNPs located more than 500 kb from the analyzed CpG sites, we identified 5,342 SNP-CpG pairs showing significant associations between genotypes and the degree of DNA methylation in adipose tissue after correction for multiple testing

Table 2. Number of significant mQTL results in human adipose tissue.

	cis-mQTL	*trans*-mQTL
SNP-CpG pairs	101,911	5,342
SNPs	51,143	2,735
CpG sites	15,208	596
Unique genes	5,589	375

Significance threshold < 0.05 after Bonferroni correction for multiple testing.
Correction value *cis* = 104,023,091
Correction value *trans* = 211,781,637,483.

doi:10.1371/journal.pone.0157776.t002

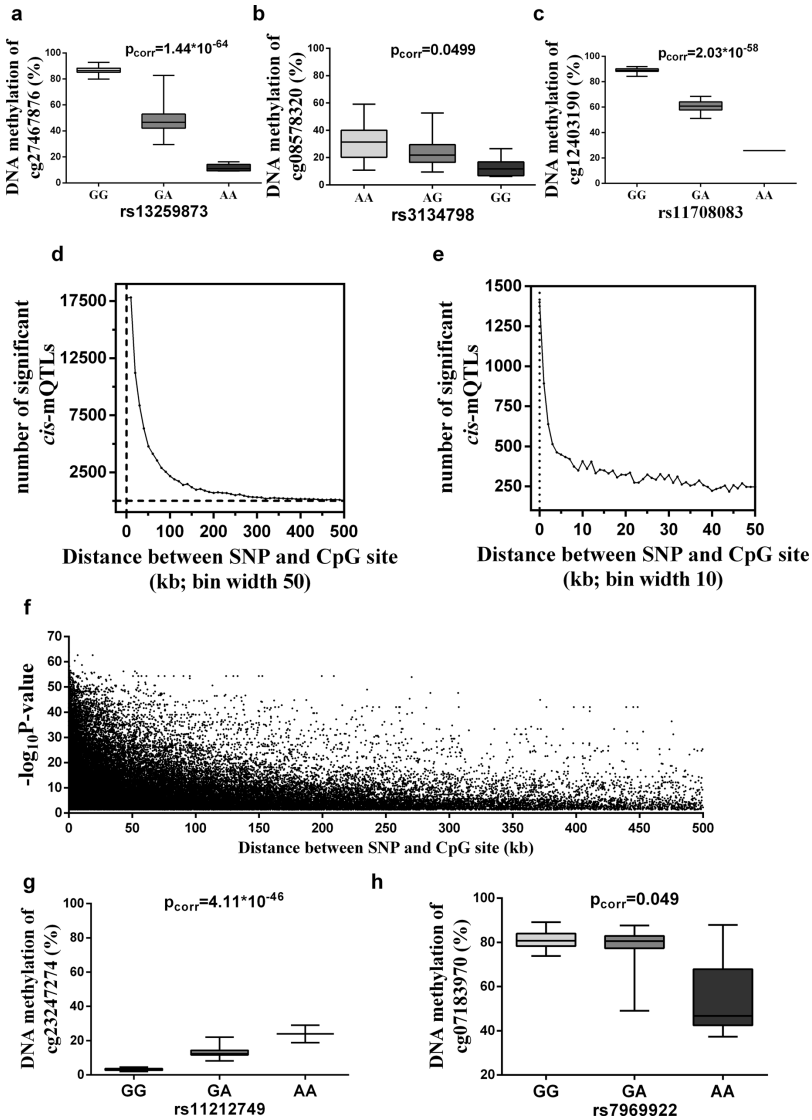

a
DNA methylation of cg27467876 (%)
p_{corr}=1.44*10^{-64}
GG
GA
AA
rs13259873
b
DNA methylation of cg08578320 (%)
p_{corr}=0.0499
AA
AG
GG
rs3134798
c
DNA methylation of cg12403190 (%)
p_{corr}=2.03*10^{-58}
GG
GA
AA
rs11708083
d
number of significant cis-mQTLs
Distance between SNP and CpG site (kb; bin width 50)
e
number of significant cis-mQTLs
Distance between SNP and CpG site (kb; bin width 10)
f
-log$_{10}$P-value
Distance between SNP and CpG site (kb)
g
DNA methylation of cg23247274 (%)
p_{corr}=4.11*10^{-46}
GG
GA
AA
rs11212749
h
DNA methylation of cg07183970 (%)
p_{corr}=0.049
GG
GA
AA
rs7969922

Fig 2. Associations between SNPs and DNA methylation in human adipose tissue. A genome-wide mQTL analysis in human adipose tissue was performed by associating SNPs with DNA methylation of CpG sites located in either *cis* (≤500 kb) or *trans*. Boxplots of (**a**) the top *cis*-mQTL, (**b**) the bottom *cis*-mQTL, and (**c**) the top *cis*-mQTL where the SNP introduces or removes a CpG site (CpG-SNP), showing significant associations between SNPs (*genotype groups*, *x-axis*) and DNA methylation of CpG sites (%, *y-axis*). (**d-e**) The frequency of associations (*y-axis*) is plotted in relation to the relative distance between SNPs and CpG sites (*kb*, *x-axis*) of significant *cis*-mQTLs. In (**d**) the full *cis*-mQTL distance of 500 kb is represented and the frequency of significant *cis*-mQTLs within each distance bin of 10kb are plotted, and, in (**e**) the region of 0-50kb is zoomed and the frequency of significant *cis*-mQTLs within in each distance bin of 1kb is plotted. (**f**) Histogram showing the strength of association (*-log$_{10}$ p-value*, *y-axis*) in relation to distance between SNP and CpG site (*kb*, *x-axis*) of significant *cis*-mQTLs. The most frequent and strongest association signals of *cis*-mQTLs are shown within SNPs located close to CpG sites. (**g-h**) Boxplots of (**g**) the top *trans*-mQTL, and (**h**) the bottom *trans*-mQTL, showing significant associations between SNPs (*genotype groups*, *x-axis*) and DNA methylation of CpG sites (%, *y-axis*). p_{corr}, p-values have been corrected for multiple testing by a modified Bonferroni correction where the LD structure of SNPs is taken into account (see methods).

doi:10.1371/journal.pone.0157776.g002

(**see** Methods), corresponding to 2,735 unique SNPs and 596 unique CpG sites (**Table 2** and **S2 Table**). Among unique CpG sites of significant *trans*-mQTLs, 366 CpG sites were annotated to 375 unique genes (**Table 2** and **S2 Table**) and 230 CpG sites were annotated to intergenic regions. The most and least significant *trans*-mQTLs are shown in **Fig 2G–2H.**

Genomic distribution of significant mQTLs in human adipose tissue

DNA methylation in proximal promoter and/or enhancer regions is generally thought to have silencing effects on gene transcription, meanwhile DNA methylation in the gene body might stimulate transcriptional elongation and contribute to alternative splicing events [6]. Giving the various functions of DNA methylation in the context of genomic regions, it is of interest to study the underlying mechanisms regulating DNA methylation patterns in different genomic regions. We therefore studied the chromosomal and genomic distribution of CpG sites in significant mQTLs in human adipose tissue. To determine whether the genomic distribution of CpG sites in significant mQTLs differ significantly from all analyzed CpG sites on the array, we performed chi-squared tests. The chromosomal distribution of CpG sites in significant *cis*- and *trans*-mQTLs is shown in **Fig 3A.** We found an overrepresentation of CpG sites in significant *cis*-mQTLs on chromosome 6, 7, 8, 13 and 21 together with an underrepresentation on chromosomes 1, 2, 3, 11, 12, 14, 15, 17, 18, 19, 20 and X when compared to the chromosomal distribution of all analyzed CpG sites (**Fig 3A**). The highest deviation from expectation of CpGs in significant *cis*-mQTLs was observed on chromosome 6 (p-value = $3.4x10^{-89}$), where the highly polymorphic HLA region is located, a genomic region linked to numerous autoimmune diseases [23,24]. CpG sites in significant *trans*-mQTLs were overrepresented on chromosomes 6 and Y while underrepresented on chromosomes 9 and 14 (**Fig 3A**).

Furthermore, the Infinium HumanMethylation450 BeadChip estimates DNA methylation in several genomic features and the analyzed CpG sites have been annotated based on their genomic location in relation to the nearest gene including genomic regions TSS1500 and TSS200 (1500–201 and 200–0 bases upstream of transcription start site (TSS), respectively), 5'UTR (untraslated region), 1st exon, gene body, 3'UTR and intergenic regions [25]. In the present study, CpG sites in significant *cis*-mQTLs were overrepresented in the intergenic regions and gene body, while significantly underrepresented in the TSS1500, TSS200, 5'UTR, 1st exon and 3'UTR (**Fig 3B**). Among significant *trans*-mQTLs, we found an overrepresentation of CpGs in the intergenic region and underrepresentation in TSS1500 and gene body (**Fig 3B**).

The analyzed CpG sites have also been annotated based on their relation to CpG islands, including the following regions: CpG islands, northern and southern shores, northern and southern shelves and open sea [25]. For CpG sites in significant *cis*-mQTLs, we found an overrepresentation in the open sea, northern- and southern shores as well as in southern shelf

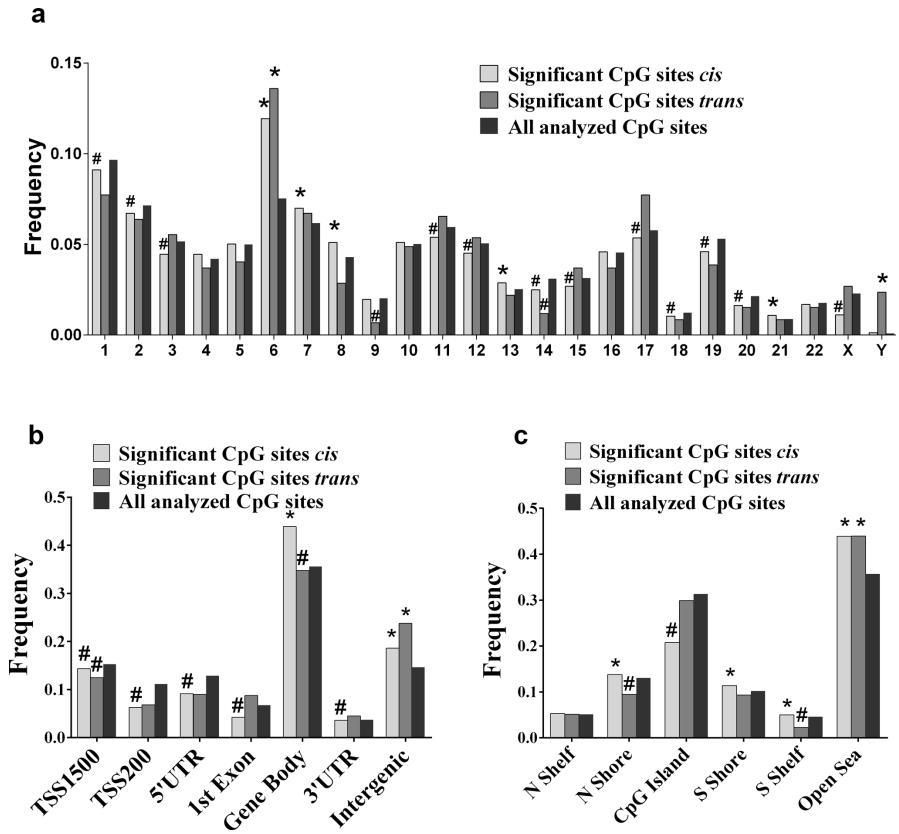


Fig 3. Distribution of CpG sites of significant mQTLs in relation to genomic regions. We examined the chromosomal and genomic distribution of CpG sites in significant mQTLs in human adipose tissue. By using chi-squared-tests, we determined whether the observed frequency of significant CpGs in *cis*- or *trans*-mQTLs differs from the frequency of all analyzed CpG sites for a particular genomic region. The histograms show the distributions of CpGs in relation to (**a**) chromosomes, (**b**) nearest gene, and (**c**) CpG islands. *Frequencies, significantly different (over-represented) from what expected by chance. #Frequencies, significantly different (under-represented) from what expected by chance. Genomic region in relation to nearest gene includes: *TSS 1500* and *TSS 200* (sites located 1500–201 or 200–0 bases upstream of the transcription start site (TSS) respectively), *5'UTR*, *1st exon*, *gene body*, *3'UTR* and *intergenic region* (not mapped to any of the other regions). Genomic region in relation to CpG island includes: *CpG island*, *shore* (flanking region of CpG island, 0–2000 bp), *shelf* (flanking region of shore, 2000–4000 bp distant from CpG island) and *open sea* (not mapped to any of the other regions).

doi:10.1371/journal.pone.0157776.g003

(Fig 3C). Moreover, an underrepresentation was found in CpG islands (Fig 3C). CpGs in significant *trans*-mQTLs showed overrepresentation in the open sea and underrepresentation in northern shore as well as southern shelf (Fig 3C).

Next, we performed a KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathway analysis to identify cellular components or biological pathways which show enrichment among

genes identified in *cis*- and *trans*- mQTL analyses in human adipose tissue. Using WebGestalt [26], we identified 172 significant (FDR < 0.05) KEGG pathways enriched among 5,589 genes annotated to significant *cis* mQTLs (**S3 Table**), including Metabolic pathways ($P_{adj} = 6.3x10^{-15}$) and Pathways in Cancer ($P_{adj} = 7.3x10^{-42}$) were found among the most enriched KEGG pathways (**S3 Table**). Moreover, 25 KEGG pathways were enriched among 375 genes annotated to significant *trans* mQTLs (**S3 Table**).

Candidate loci for obesity and diabetes related traits are detected among mQTLs in human adipose tissue

Numerous SNPs associated with obesity, type 2 diabetes and related traits have previously been identified by GWAS [1]. However, the molecular mechanisms explaining how most of these SNPs affect gene function and disease pathology remain scarce. We therefore tested if identified SNPs in significant mQTLs in adipose tissue overlap with loci previously reported to associate with obesity, type 2 diabetes or obesity/diabetes related traits in the GWAS catalog ($p<10^{-5}$) [27]. Out of the SNPs significantly associated with DNA methylation in the *cis*-mQTL analysis and when taking proxy SNPs into account ($R^2>0.8$, **see** Methods), 19,706 overall, we found 231 SNPs of significant mQTLs that overlapped with at least one of the 2138 reported disease or trait locus identified in the GWAS catalog (**S4 Table**), which constitutes 1.17% of *cis*-mQTL SNPs and 10.8% of GWAS catalog SNPs. Representative mQTLs for some of these loci are shown in **Fig 4A–4J**. These mQTLs include *POMC* and *LEPR*, which encode proopiomelanocortin and the leptin receptor, respectively. Mutations in both these genes have been associated with early onset obesity [28]. We also present mQTLs covering *GIPR* (encoding gastric inhibitory polypeptide receptor), *PARP4* (encoding poly(ADP-ribosyl)transferase-like 1 protein), *CEPT* (encoding cholesteryl ester transfer protein), *APOA5* (encoding apolipoprotein A5), *SORT1* (encoding sortilin 1), *GCKR* (encoding glucokinase regulator), *FADS2* (encoding fatty acid desaturase 2), *ACADS* (encoding acyl-CoA dehydrogenase) and *GRB10* (encoding growth factor receptor bound protein 10). SNPs in these loci have previously been associated with BMI, T2D and/or obesity- and lipid-related traits [29–36].

Of SNPs in significant *trans*-mQTLs, we found 4 SNPs overlapping with reported obesity loci in the GWAS catalog (**Fig 4K and S5 Table**).

The impact of identified mQTLs on mRNA expression in human adipose tissue

It is well established that mRNA expression is regulated by both genetic variation and DNA methylation independently [4,37]. However, the insights of how genetic and epigenetic variation interacts to influence gene expression remain limited. In order to study the impact of identified mQTLs on mRNA expression in human adipose tissue, we performed a follow-up eQTL analysis in 118 samples with available microarray expression data (out of original 119 samples). First, we related the 51,143 unique SNPs, showing significant association with DNA methylation in the *cis*-mQTL analysis, with mRNA expression of genes within 500kb (*cis*-distance). In the eQTL analysis of significant *cis*-mQTL SNPs, we identified 926 SNP-mRNA transcript pairs showing significant associations between genotypes and mRNA expression levels after correction for multiple testing (**see** Methods). These correspond to 635 unique SNPs and 86 unique genes, including *CHRNA5*, *G6PC2*, *GPX7*, *RPL27A*, *THNSL2* and *ZFP57* (**Table 3, Fig 5 and S6 Table**). *CHRNA5* encodes a nicotinic acetylcholine receptor subunit and SNPs in this locus have been associated with body weight in relation to tobacco use [38]. *G6PC2* encodes glucose-6-phosphatase catalytic subunit 2 and SNPs in this locus have been associated with glycemic traits [39]. *GPX7* encodes glutathione peroxidase 7 a protein involved in glutathione

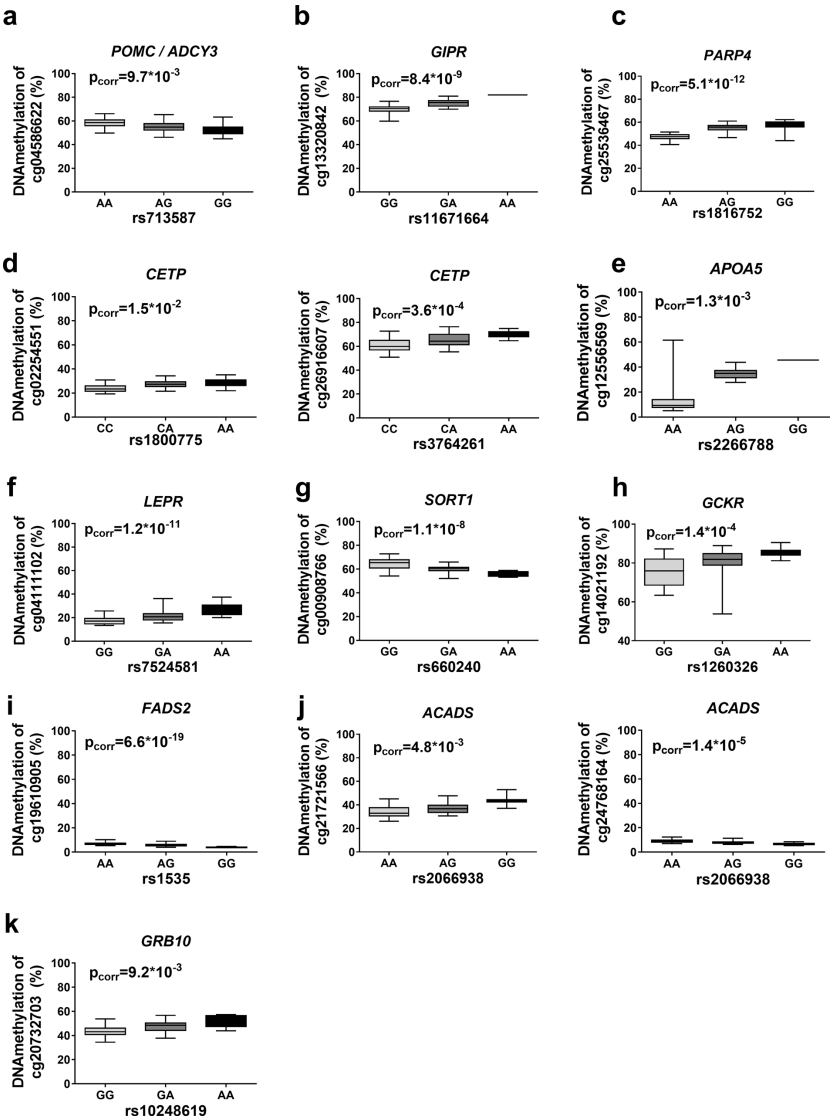

a
POMC / ADCY3
DNAmethylation of cg04586622 (%)
p_corr=9.7*10^-3
AA AG GG
rs713587
b
GIPR
DNAmethylation of cg13320842 (%)
p_corr=8.4*10^-9
GG GA AA
rs11671664
c
PARP4
DNAmethylation of cg25536467 (%)
p_corr=5.1*10^-12
AA AG GG
rs1816752
d
CETP
DNAmethylation of cg02254551 (%)
p_corr=1.5*10^-2
CC CA AA
rs1800775
CETP
DNAmethylation of cg26916607 (%)
p_corr=3.6*10^-4
CC CA AA
rs3764261
e
APOA5
DNAmethylation of cg12556569 (%)
p_corr=1.3*10^-3
AA AG GG
rs2266788
f
LEPR
DNAmethylation of cg04111102 (%)
p_corr=1.2*10^-11
GG GA AA
rs7524581
g
SORT1
DNAmethylation of cg00908766 (%)
p_corr=1.1*10^-8
GG GA AA
rs660240
h
GCKR
DNAmethylation of cg14021192 (%)
p_corr=1.4*10^-4
GG GA AA
rs1260326
i
FADS2
DNAmethylation of cg19610905 (%)
p_corr=6.6*10^-19
AA AG GG
rs1535
j
ACADS
DNAmethylation of cg21721566 (%)
p_corr=4.8*10^-3
AA AG GG
rs2066938
ACADS
DNAmethylation of cg24768164 (%)
p_corr=1.4*10^-5
AA AG GG
rs2066938
k
GRB10
DNAmethylation of cg20732703 (%)
p_corr=9.2*10^-3
GG GA AA
rs10248619

Fig 4. mQTLs in adipose tissue capture reported disease loci. Depiction of some identified mQTLs in adipose tissue of previously reported GWAS loci associated with obesity: (**a**) *POMC* / *ADCY3*, (**b**) *GIRP*, and (**c**) *PARP4*; lipid profiles, waist and metabolic syndrome: (**d**) *CETP*, (**e**) *APOA5*, (**f**) *LEPR*, (**g**) *SORT1*, (**h**) *GCKR* and (**i**) *FADS2*; and metabolic traits: (**j**) *ACADS* and (**k**) *GRB10*. *ADCY3* locus and *LEPR* loci were identified through proxy SNPs based on LD.

doi:10.1371/journal.pone.0157776.g004

metabolism. *RPL27A* encodes Ribosomal protein L27A, which has been linked to human obesity [40]. *THNSL2* encodes threonine synthase like 2 and SNPs in this locus have been associated with obesity [41]. Additionally, *ZFP57* encodes a zink finger protein and DNA methylation and mutations in this locus are associated with transient neonatal diabetes [42].

The 2,735 unique SNPs identified in the *trans*-mQTL analysis were also followed-up and related to mRNA expression of all analyzed genes. In the eQTL analysis of significant *trans*-mQTL SNPs, we identified 270 significant associations between genotypes and mRNA expression levels after correction for multiple testing (**see** Method), consisting of 89 unique SNPs and 10 unique genes *e.g. GSTT1*, *HLA-DQB1* and *ZFP57* (**Table 3 and S7 Table**).

The impact of identified mQTLs on metabolic phenotypes

Given that adipose tissue contributes to whole body energy homeostasis by for instance insulin-stimulated glucose uptake, triglyceride storage and adipokine secretion, we investigated if the identified SNPs in significant mQTLs affect metabolic phenotypes in our study cohort. Identified mQTL SNPs were related to obesity measurements, glycemic traits and lipid levels in our study cohort of 119 Scandinavian men as well as looked-up in public available consortia data from the GIANT [43,44], MAGIC [36,45,46] and GLGC [47] consortia. Out of the significant *cis*-mQTLs, we found 62 SNPs associated with BMI, 185 with waist-hip ratio (WHR), 77 with fasting glucose, 62 with fasting insulin, 91 with homeostasis model of beta-cell function (HOMA-B), 49 with HOMA-IR, 146 with HbA1c, 85 with total cholesterol, 84 with triglycerides, 197 with HDL, 67 with LDL in both our study cohort and consortia data with the same direction of allele effects and with $P \leq 0.05$ (**S8 Table**). Several of these SNPs show genome-wide significance in GIANT, MAGIC or GLGC. Representative associations between genotype and some metabolic traits as well as DNA methylation in the 119 Scandinavian men are shown in **Fig 6A–6C.** The SNPs presented in **Fig 6** do also show genome-wide significance with respective trait in GLGC (*rs2523453*, cholesterol, $p = 6.5^{*}10^{-08}$ and *rs7205804*, HDL, $p = 5.27^{-675}$) and MAGIC (*rs11603334*, fasting glucose, $p = 2.9^{*}10^{-08}$), respectively (**S8 Table**).

Table 3. Number of significant eQTL results in human adipose tissue.

	eQTLs of *cis*-mQTL-SNPs	eQTLs of *trans*-mQTL-SNPs
SNP-mRNA transcript pairs	926	270
Unique SNPs	635	89
Unique mRNA transcripts	101	14
Unique genes	86	10

Only SNPs of significant mQTLs are included in the eQTL analysis.
SNPs of significant *cis*-mQTLs are regressed against mRNA expression of mRNA transcripts located in *cis* (≤ 500kb).
SNPs of significant *trans*-mQTLs are regressed against mRNA expression of all mRNA transcripts.
Significance threshold < 0.05 after correction for multiple testing.
Correction value for eQTL analysis for *cis*-mQTL-SNPs = 934,021
Correction value for eQTL analysis for *trans*-mQTL-SNPs = 33,326,082.

doi:10.1371/journal.pone.0157776.t003

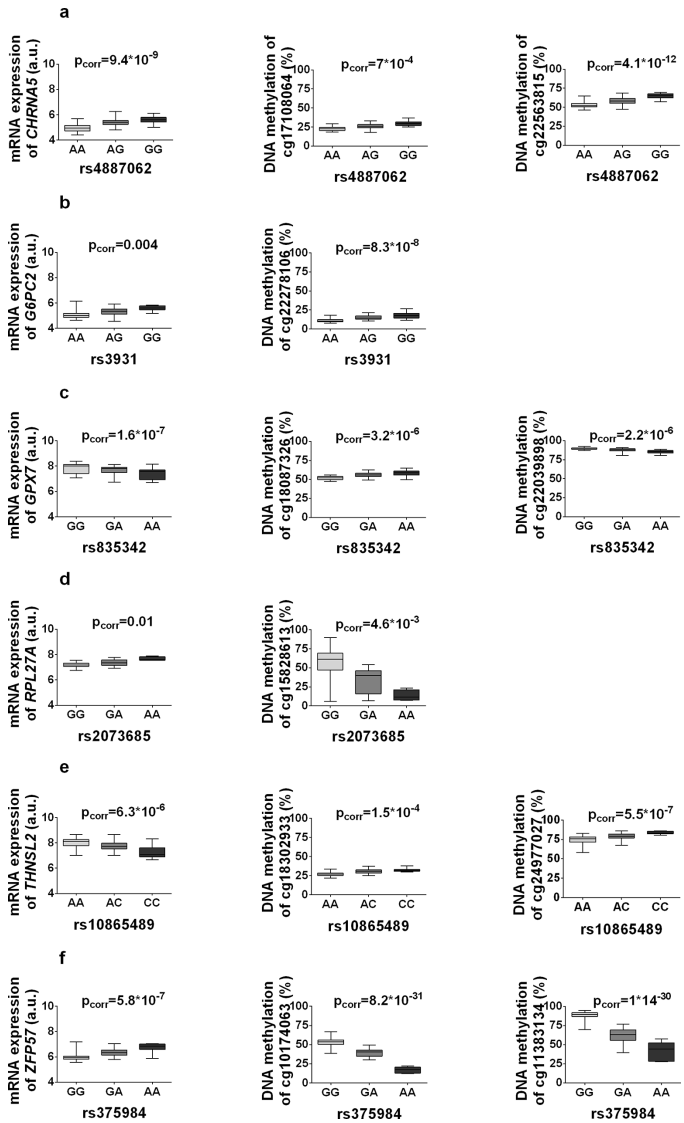

a
mRNA expression of CHRNA5 (a.u.)
pcorr=9.4*10-9
AA AG GG
rs4887062
DNA methylation of cg17108064 (%)
pcorr=7*10-4
AA AG GG
rs4887062
DNA methylation of cg22563815 (%)
pcorr=4.1*10-12
AA AG GG
rs4887062
b
mRNA expression of G6PC2 (a.u.)
pcorr=0.004
AA AG GG
rs3931
DNA methylation of cg22278106 (%)
pcorr=8.3*10-8
AA AG GG
rs3931
c
mRNA expression of GPX7 (a.u.)
pcorr=1.6*10-7
GG GA AA
rs835342
DNA methylation of cg18087326 (%)
pcorr=3.2*10-6
GG GA AA
rs835342
DNA methylation of cg22039898 (%)
pcorr=2.2*10-6
GG GA AA
rs835342
d
mRNA expression of RPL27A (a.u.)
pcorr=0.01
GG GA AA
rs2073685
DNA methylation of cg15828613 (%)
pcorr=4.6*10-3
GG GA AA
rs2073685
e
mRNA expression of THNSL2 (a.u.)
pcorr=6.3*10-6
AA AC CC
rs10865489
DNA methylation of cg18302933 (%)
pcorr=1.5*10-4
AA AC CC
rs10865489
DNA methylation of cg24977027 (%)
pcorr=5.5*10-7
AA AC CC
rs10865489
f
mRNA expression of ZFP57 (a.u.)
pcorr=5.8*10-7
GG GA AA
rs375984
DNA methylation of cg10174063 (%)
pcorr=8.2*10-31
GG GA AA
rs375984
DNA methylation of cg11383134 (%)
pcorr=1*14-30
GG GA AA
rs375984

Fig 5. mQTLs affect gene expression in human adipose tissue. Significant mQTL SNP-CpG pairs where the SNP also shows significant association with gene expression in adipose tissue. The boxplots represent some identified mQTL SNPs and associations of the same loci with mRNA expression: (**a**) *CHRNA5*, (**b**) *G6PC2*, **(c)** *GPX7*, **(d)** *RPL27A*, **(e)** *THNSL2* and **(f)** *ZFP57*. Annotations for these mQTLs are included in S1 Table.

doi:10.1371/journal.pone.0157776.g005

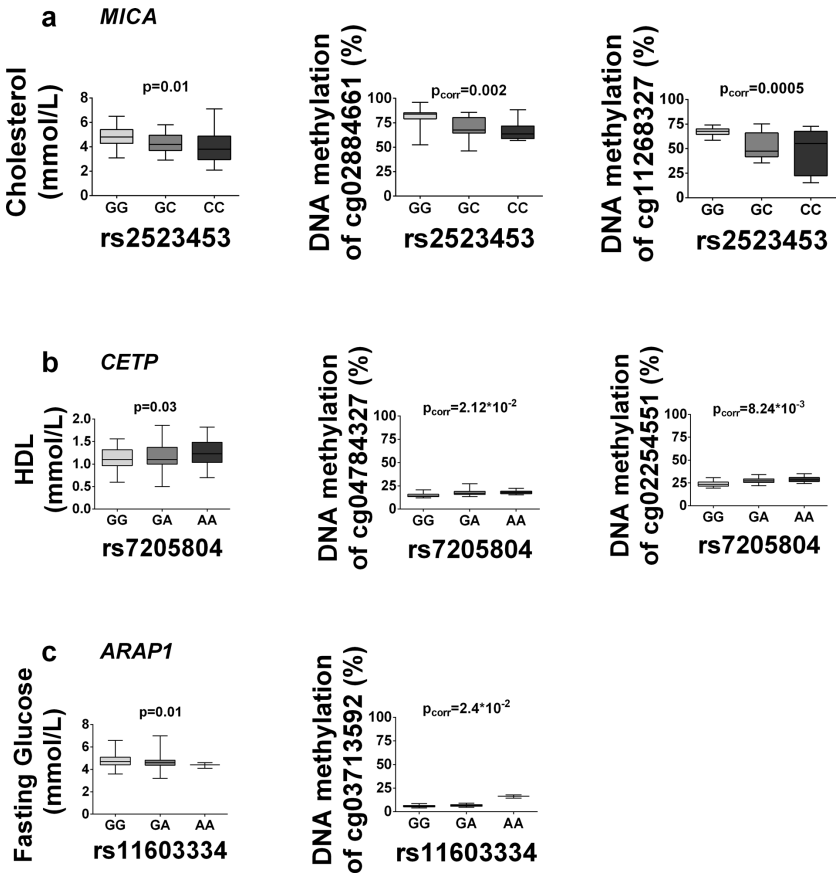


Fig 6. mQTLs in human adipose tissue affect metabolic phenotypes. The boxplots show significant mQTL SNPs associated with metabolic phenotypes in our study cohort with p<0.05, and associations of these loci with DNA methylation in adipose tissue for (**a**) *rs2523453*, (**b**) *rs7205804*, **(c)** *rs11603334*.

doi:10.1371/journal.pone.0157776.g006

Additionally, 8 loci detected in the overlap of the *cis*-mQTL and eQTL analysis were among those associated with metabolic phenotypes (**S8** and **S9 Tables**). These data show the effect of interactions between common genetic variation and DNA methylation on gene expression and metabolic outcome (depiction presented in **Fig 7A–7E**, all overlapping SNPs presented in **S9 Table**).

Of identified *trans*-mQTLs, we found 2 SNPs associated with BMI, 13 with WHR ratio, 6 with fasting glucose, 1 with fasting insulin, 2 with HOMA-IR, 42 with HbA1c, 6 with total cholesterol, 7 with triglycerides, 68 with HDL, and 4 with LDL in both our study cohort and consortia data with the same direction of allele effects and with P $\leq$0.05 (**S10 Table**).

Additionally, 2 of the identified *cis*-mQTL SNPS were previously found to be associated with C-reactive protein (CRP) levels (S11 Table).

Additionally, some of the identified *cis*- and *trans*- mQTLs are annotated to candidate genes for adipose-related traits. Out of the 157 loci previously implicated in lipid biology in GLGC consortium [47], 48 (30%) were found among 5,589 unique genes annotated to significant cis-mQTLs (S1 Table), and 4 among 375 unique genes annotated to significant trans-mQTLs (S2 Table).

Causality inference test (CIT)–DNA methylation potentially mediates the genetic impact on metabolic phenotypes

We proceeded to evaluate the potential causality relationships between genotypes (G), DNA methylation (M) and phenotypic traits (P) using the CIT [21]. The possible relationships between these three factors are shown in **Fig 8.** The CIT was performed in our cohort of 119 Scandinavian men for identified SNP-CpG pairs in the mQTL analysis where the SNP also showed significant association with a metabolic phenotype in both our study cohort and publicly available consortia data with P $\leq$0.05. For *cis*-mQTLs, we identified 39 SNP-CpG pairs, corresponding to 35 unique SNPs and 22 unique CpGs, where SNP plays a causal role on metabolic phenotype, mediated by DNA methylation (**Table 4**). Out of these 39 SNP-CpG pairs, 1 pair was significantly associated with BMI, 2 for fasting glucose, 1 for fasting insulin, 1 for HOMA-B, 7 for HOMA-IR, 7 for HbA1c, 9 for cholesterol, 1 for triglycerides 5 for HDL and 5 for LDL (**Table 4**). Among the genes annotated to these SNP-CpG pairs, *CDK2AP1*, *HLA-DMA*, *MCM6*, *TCF19*, *CAMK1D* and *NEIL2* were found. None of the *cis*-mQTLs showed a reactive relationship between a SNP and a metabolic phenotype.

Biological replication of mQTLs in human adipose tissue

To validate whether the results of mQTL analysis hold in an independent cohort, we also looked for overlap with a recent study also showing associations between genetic variation and DNA methylation in human adipose tissue [16]. While both studies analyzed DNA methylation using the Infinium HumanMethylation450 BeadChip, Grundberg *et al*. restricted their mQTL analysis to SNPs located within 100 kb from analyzed CpG sites [16] and therefore it was only possible to compare some of our results. It should also be taken into account that while our study included men, the study by Grundberg *et al*. included women and the two studies used different bioinformatic and statistical approaches, which may affect the possibility to replicate the results. Nevertheless, among our significant *cis*-mQTLs, we found that 5,468 CpG sites also stand under genetic control of SNPs in the study by Grundberg *et al*., and out of these 2,118 (38.6%) were associated with the same SNP in both their and our study [16].

Additionally, we recently performed an mQTL analysis in human pancreatic islets [20]. Here, we looked for overlap between the significant mQTLs identified in human adipose tissue of the 119 men and the mQTLs previously found in human pancreatic islets [20]. Among our

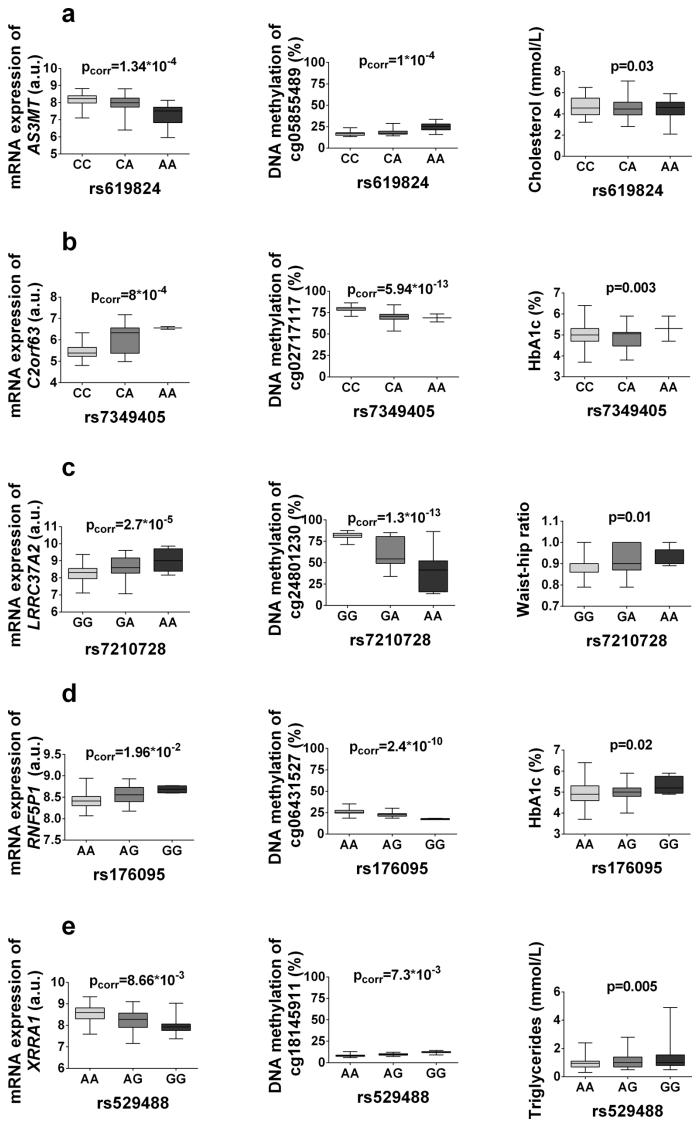

a
mRNA expression of AS3MT (a.u.)
pcorr=1.34*10-4
CC CA AA
rs619824
DNA methylation of cg05855489 (%)
pcorr=1*10-4
CC CA AA
rs619824
Cholesterol (mmol/L)
p=0.03
CC CA AA
rs619824
b
mRNA expression of C2orf63 (a.u.)
pcorr=8*10-4
CC CA AA
rs7349405
DNA methylation of cg02717117 (%)
pcorr=5.94*10-13
CC CA AA
rs7349405
HbA1c (%)
p=0.003
CC CA AA
rs7349405
c
mRNA expression of LRRC37A2 (a.u.)
pcorr=2.7*10-5
GG GA AA
rs7210728
DNA methylation of cg24801230 (%)
pcorr=1.3*10-13
GG GA AA
rs7210728
Waist-hip ratio
p=0.01
GG GA AA
rs7210728
d
mRNA expression of RNF5P1 (a.u.)
pcorr=1.96*10-2
AA AG GG
rs176095
DNA methylation of cg06431527 (%)
pcorr=2.4*10-10
AA AG GG
rs176095
HbA1c (%)
p=0.02
AA AG GG
rs176095
e
mRNA expression of XRRA1 (a.u.)
pcorr=8.66*10-3
AA AG GG
rs529488
DNA methylation of cg18145911 (%)
pcorr=7.3*10-3
AA AG GG
rs529488
Triglycerides (mmol/L)
p=0.005
AA AG GG
rs529488

Fig 7. mQTLs/eQTLs in human adipose tissue affect metabolic phenotypes. Significant mQTL SNPs associated with both gene expression and a metabolic phenotype, with boxplots showing associations of some of these loci with DNA methylation, gene expression and metabolic traits: (**a**) *rs619824*, (**b**) *rs7349405*, (**c**) *rs7210728*, (**d**) *rs176095*, (**e**) *rs529488*.

doi:10.1371/journal.pone.0157776.g007

significant *cis*-mQTLs in adipose tissue, 39,386 were also found in pancreatic islets (**S12 Table**). Moreover, 1,852 significant *trans*-mQTLs overlapped between the two different tissues (**S13 Table**).

mQTLs in human adipose tissue do also show differential DNA methylation in patients with type 2 diabetes

We have previously identified CpG sites that are differentially methylated in adipose tissue from subjects with type 2 diabetes compared with non-diabetic controls [15]. However, it remains unknown if methylation of these sites may also be under genetic control. Therefore, we further tested if these CpG sites [15] overlap with our significant *cis* and *trans* mQTLs in human adipose tissue (**S1 and S2 Tables**). Interestingly, we discovered that 237 CpG sites among our significant *cis*-mQTLs and 7 CpG sites among our significant *trans*-mQTLs are also differentially methylated in adipose tissue from subjects with type 2 diabetes (**S14 Table**), suggesting that DNA methylation may mediate the genetic impact of type 2 diabetes.

mQTLs in human adipose tissue overlap with CpG sites associated with BMI and HbA1c

We have previously identified CpG sites for which the adipose tissue methylation level associates with BMI and HbA1c [48]. Here, we examined if these CpG sites overlap with our *cis* and *trans*-mQTLs in human adipose tissue (**S1 and S2 Tables**). We found that 33,058 CpG sites previously identified as associated with BMI overlapped with 577 *cis* and 19 *trans* significant mQTLs in current study (**S15 Table**). Moreover, out of 711 CpG sites associated with HbA1c, 25 and 1 CpG site overlapped with significant *cis* and *trans* mQTLs respectively (**S15 Table**).

mQTL analyses in adipose tissue of two sub-cohorts

Since the subjects in the four sub-cohorts included in this study differ in age and BMI, we performed a sub-analysis only including cohorts #1 and #2 as these subjects are phenotypically similar. Here, we detected 66,329 mQTLs in *cis* showing significant associations between genotype and the degree of DNA methylation after correction for multiple testing, corresponding to

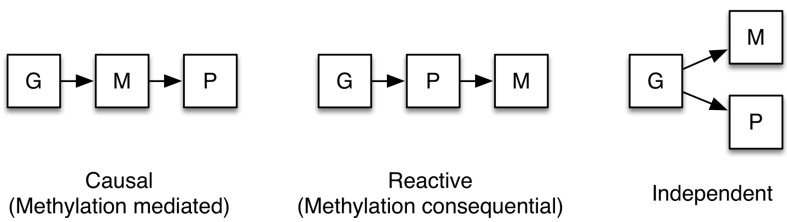


Fig 8. CIT in human adipose tissue. Possible relationship models between genotype as a causal factor (G), DNA methylation as the mediator factor (M) and metabolic phenotype as the phenotypic outcome (P).

doi:10.1371/journal.pone.0157776.g008

Table 4. Identified *cis*-mQTLs where DNA methylation potentially mediates the interactions between a genotype and a phenotype in human adipose tissue.

Chr	CpG Id	CpG Gene	CpG Gene Region	SNP Id	SNP Gene	G vs M p_{corr}-value	Phenotype	G vs P p-value	CIT causal p-value
6	cg12929486	*SLC22A16*	TSS200	*rs2428190*	*SLC22A16*	1.30E-07	BMI	0.049	0.02
5	cg14825688	*LEAP2*	TSS1500	*rs39830*	*UQCRQ*	4.02E-06	Fasting glucose	0.015	0.03
5	cg14825688	*LEAP2*	TSS1500	*rs803217*	-	5.20E-05	Fasting glucose	0.018	0.03
2	cg01726273	-	Intergenic	*rs4853438*	*SNRPG*	3.60E-02	Fasting insulin	0.050	0.03
8	cg11123440		Intergenic	*rs12458*	*GATA4*	8.99E-03	HOMA-B	0.042	0.03
12	cg10240950	*C12orf76*	Body	*rs1027949*	*GIT2*	3.91E-02	HOMA-IR	0.036	0.05
12	cg10240950	*C12orf76*	Body	*rs10774978*	*TCHP*	3.91E-02	HOMA-IR	0.030	0.05
12	cg10240950	*C12orf76*	Body	*rs11068984*	*GIT2*	3.91E-02	HOMA-IR	0.017	0.05
10	cg26169081	*CAMK1D; CAMK1D*	Body;Body	*rs11257926*	*CAMK1D*	1.33E-03	HOMA-IR	0.021	0.04
10	cg26169081	*CAMK1D; CAMK1D*	Body;Body	*rs17152029*	*CAMK1D*	3.92E-04	HOMA-IR	0.003	0.01
10	cg26169081	*CAMK1D; CAMK1D*	Body;Body	*rs17152037*	*CAMK1D*	1.33E-03	HOMA-IR	0.005	0.04
12	cg10240950	*C12orf76*	Body	*rs2302689*	*ANKRD13A*	3.91E-02	HOMA-IR	0.030	0.05
7	cg17372657		Intergenic	*rs1880296*	-	2.69E-07	HbA1c	0.032	0.03
7	cg17372657		Intergenic	*rs2949170*	-	1.79E-04	HbA1c	0.034	0.03
7	cg17372657		Intergenic	*rs2949192*	-	9.02E-05	HbA1c	0.014	0.03
6	cg13561028	*SFTA2*	Body	*rs3130782*	*LOC100129065*	3.68E-09	HbA1c	0.002	0.01
6	cg13561028	*SFTA2*	Body	*rs3131934*	-	8.49E-08	HbA1c	0.002	0.04
16	cg04544033	-	Intergenic	*rs556179*	-	2.37E-02	HbA1c	0.014	0.03
6	cg13561028	*SFTA2*	Body	*rs7750641*	*TCF19*	1.27E-17	HbA1c	0.028	0.02
12	cg21745287	*ARL6IP4; OGFOD2*	TSS1500;3'UTR	*rs10846489*	*CDK2AP1*	4.91E-03	Cholesterol	0.013	0.04
12	cg07644039	*ARL6IP4; OGFOD2*	TSS1500;3'UTR	*rs10846489*	*CDK2AP1*	2.34E-02	Cholesterol	0.013	0.04
12	cg21745287	*ARL6IP4; OGFOD2*	TSS1500;3'UTR	*rs1109559*	-	4.84E-03	Cholesterol	0.011	0.03
12	cg07644039	*ARL6IP4; OGFOD2*	TSS1500;3'UTR	*rs1109559*	-	1.44E-02	Cholesterol	0.011	0.03
12	cg07644039	*ARL6IP4; OGFOD2*	TSS1500	*rs4275659*	*ABCB9*	2.24E-03	Cholesterol	0.010	0.02
12	cg21745287	*ARL6IP4; OGFOD2*	TSS1500;3'UTR	*rs6488868*	*SBNO1*	9.39E-04	Cholesterol	0.008	0.02
15	cg12371991	-	Intergenic	*rs6494591*	-	8.73E-03	Cholesterol	0.038	0.01
2	cg04490207	-	Intergenic	*rs6712567*	-	1.15E-02	Cholesterol	0.034	0.03
9	cg14341289	*FSD1L*	TSS1500	*rs885954*	-	1.81E-02	Cholesterol	0.041	0.00
8	cg00820056	-	Intergenic	*rs11787024*	*LY6H*	6.37E-03	Triglycerides	0.033	0.04
6	cg14833385	*HLA-DMA*	TSS1500	*rs1480380*	-	3.50E-17	HDL	0.007	0.03
2	cg01726273	-	Intergenic	*rs2921711*	*TIA1*	3.60E-02	HDL	0.037	0.04
2	cg07169764	*MCM6;MCM6*	1stExon;5'UTR	*rs309172*	*DARS*	4.06E-07	HDL	0.011	0.01
8	cg05875700	*ERICH1*	Body	*rs3735917*	*ERICH1*	3.17E-11	HDL	0.008	0.05
2	cg07169764	*MCM6;MCM6*	1stExon;5'UTR	*rs6750549*	*DARS*	4.06E-07	HDL	0.010	0.01
4	cg08029340	*MYL5*	Body	*rs11726338*	*PIGG*	9.63E-03	LDL	0.028	0.04
6	cg02525939	-	Intergenic	*rs4710698*	-	6.17E-07	LDL	0.007	0.04
6	cg04399728	-	Intergenic	*rs4710698*	-	2.69E-13	LDL	0.007	0.04

(*Continued*)

Table 4. (*Continued*)

Chr	CpG Id	CpG Gene	CpG Gene Region	SNP Id	SNP Gene	G vs M p_{corr}-value	Phenotype	G vs P p-value	CIT causal p-value
9	cg14341289	*FSD1L*	TSS1500	*rs885954*	-	1.81E-02	LDL	0.005	0.00
2	cg09644356	-	Intergenic	*rs940670*	-	2.94E-06	LDL	0.001	0.03

doi:10.1371/journal.pone.0157776.t004

36,909 unique SNPs and 11,788 unique CpG sites (**S16 Table**). Out of those 66,329 mQTLs, 63,714 (96%) overlapped with the analysis of all 4 cohorts.

In the *trans*-mQTL analysis, we identified 3,243 SNP-CpG pairs showing significant associations between genotypes and the degree of DNA methylation in adipose tissue after correction for multiple testing, corresponding to 1,865 unique SNPs and 538 unique CpG sites (**S17 Table**). Out of those 3,243 mQTLs, 2,919 (90%) were previously identified in the analysis of all 4 cohorts.

mQTL analyses in adipose tissue without adjusting for BMI

In order to validate whether BMI as a covariate has a significant effect on a number of discovered mQTLs, we performed a mQTL analysis of all 4 study cohorts without BMI as a covariate. Overall, we detected 102,467 significant *cis* mQTLs corresponding to 51,435 unique SNPs and 15,267 unique CpG sites. Out of those, 99,661 (97.2%) were also identified in the analysis where BMI was included as a covariate (**S18 Table**). In *trans*, we discovered 5,435 significant mQTLs, corresponding to 608 unique CpG sites and 2,765 unique SNPs, where 5,272 (97%) were also identified in the main mQTL analysis (**S19 Table**).

Associations between DNA methylation and mRNA expression in human adipose tissue

We finally tested the direct association between DNA methylation and gene expression in human adipose tissue by performing a linear regression between individual mRNA transcripts and DNA methylation of CpG sites in cis (500 kb up- and 100 kb downstream of respective gene) including age, BMI and study cohort as covariates. We found significant associations between DNA methylation and mRNA expression for 546 combinations (FDR<5%), consisting of 473 unique CpG sites and 194 unique mRNA transcripts (**S20 Table**), which are annotated to 173 genes.

In addition, we found that 262 CpG sites among our significant cis-mQTLs and 13 among our significant trans-mQTLs overlapped with methylation sites associated with mRNA expression in adipose tissue (**S20 Table**).

Discussion

The present study highlights the importance of genome-wide interactions between genetic and epigenetic variation and its role in human metabolism. Using CIT tests, we could for the first time identify adipose tissue methylation-mediated relationships between genotype and metabolic phenotypes, including lipid and glucose traits. Importantly, these data demonstrate how genetic variants may mediate their effects on metabolic traits via altered DNA methylation in human adipose tissue. Additionally, numerous identified mQTL-SNPs cover previously identified GWAS loci for obesity, lipid and diabetes related traits e.g. *POMC*, *GIPR*, *GRB10*, *FADS2*, *SORT1* and *APOA5*.

Multiple SNPs identified through GWAS associate with complex metabolic disease including obesity and type 2 diabetes [29,31,43,49–52]. However, the effect sizes of common variants influencing these diseases are often modest and in total they only explain small proportions of the estimated genetic predispositions to the diseases. Epigenetic factors such as DNA methylation have also been shown to be involved in the pathogenesis of various metabolic diseases [7,9,15,29,53–60]. However, studies examining the genetic regulation of inter-individual variation in DNA methylation and its contribution to metabolic outcomes are scarce but would likely give new insights to the field. Here, we performed a genome-wide mQTL analysis looking at both *cis* and *trans* effects of genetic variation on DNA methylation in human adipose tissue. To further link identified mQTLs with biological functions, we performed follow-up analyses of significant mQTL SNPs with gene expression in adipose tissue and metabolic phenotypes in our study cohort. We also looked for overlap with disease loci reported to associate with obesity and diabetes related traits in GWAS. All together, we found 101,911 SNP-CpG pairs in *cis* and 5,342 SNP-CpG pairs in *trans* showing significant associations between genotype and DNA methylation in adipose tissue demonstrating a strong genetic impact on DNA methylation in human adipose tissue. Our data are in line with previous mQTL analyses, which also show strong interactions between genetic and epigenetic variation [16–20,61], and in concordance, we found an enrichment of *cis*-mQTLs in a short distance window between associated SNPs and CpG sites. However, while most previous mQTLs have been limited to studying promoter regions [17–19] or *cis* interactions [16], we can for the first time present mQTL results in adipose tissue looking at both *cis* and *trans* effects in most genomic regions and genes. Interestingly, we observe a higher than expected number of methylation sites in significant mQTLs located in intergenic regions, in the gene body and outside of CpG islands. This observation is in line with previous studies showing that differentially methylated sites in response to environmental or genetic factors to a higher extent than expected are located outside CpG islands or within intergenic and gene body regions [10,16]. It may be that promoter regions are rich in CpG islands which are hypomethylated and are more evolutionary conserved based on their biological function, meanwhile non-CpG islands are more methylated and dynamic [6,62–64]. Interestingly, we demonstrate for the first time an enrichment of significant mQTLs in adipose tissue on chromosome 6. This chromosome possesses a highly polymorphic gene region coding the HLA complexes which are known to be implicated in several autoimmune disorders and inflammation processes [23,24]. Numerous loci identified in the *cis*- and *trans*-mQTL analysis, as well as genes in the eQTL follow-up analysis, are linked to the *HLA* genes. Based on this finding, we investigated the link between mQTLs on chromosome 6 and a measure of inflammation e.g. i.e. CRP levels. Interestingly, we found that 2 SNPs in significant mQTLs cover GWAS loci associated with CRP levels (**S11 Table**). However, none of them was located on chromosome 6 [65].

Genetic association studies have improved our understanding of the biological basis of metabolic disease [66]. Nevertheless, the effect of numerous reported obesity and diabetes SNPs on target genes or biology still remains unknown. Investigating the genetic control of variation in DNA methylation may improve our understanding of biological processes and linking loci to tissue dependent phenotypes and diseases. Elevating, we found that several SNPs associated with DNA methylation show impact on metabolic phenotypes in the studied cohort, including obesity measurements, glucose- and insulin traits, as well as lipid profiles. The effect of mQTLs on molecular phenotypes was further supported by independent replication in consortia data of obesity measurements from GIANT [43,44], glucose traits from MAGIC [36,45,46] and lipid profiles from GLGC [47]. Although mQTL SNPs were only showing nominal association to metabolic phenotypes in our study cohort, the overlap and replication in independent studies, based on consortium data, support effects of these SNPs on biological function. Indeed,

several of these SNPs show genome-wide significance in previous GWAS [47,66–68]. These include SNPs associated with cholesterol levels and annotated to *ANKRD31* (ankyrin repeat domain 31), HDL levels and annotated to *CELSR2* (cadherin, EGF LAG seven-pass G-type receptor 2) as well as fasting plasma glucose levels and annotated to *ARAP1* (ankyrin repeat and PH domain 1).

Given that SNPs affect DNA methylation and that DNA methylation is a dynamic process that may change in response to environmental factors and affects phenotype transmission [10,69], it may be possible that the SNP effect on DNA methylation levels, and indirectly on metabolic phenotypes, may change under different environmental conditions. It is hence possible that some of the identified mQTL SNPs overlapping with consortium data may have escaped detection to disease phenotypes in previous GWAS studies since DNA methylation levels was not considered. This form of gene-environment interactions could potentially affect the SNPs impact on disease risk. Indeed, our previous data, where we identified a SNP that introduces a CpG site in the promoter of *NDUFB6*, support this hypothesis [60]. Here, we showed that while elderly carriers of the genotype that introduces a CpG site had a high degree of methylation in the SNP-CpG site together with decreased skeletal muscle *NDUFB6* expression and decreased glucose uptake, young carriers had low degree of methylation in the SNP-CpG site together with increased skeletal muscle *NDUFB6* expression and no effect on glucose uptake. Together, this study demonstrates a clear interaction between genetic, epigenetic and non-genetic factors. Additionally, genetic variation may carry inheritance of epigenetic variation and thereby have an impact on the heritability of human diseases and may explain some of the missing heritability of human complex diseases. Furthermore, we also found that several SNPs associated with DNA methylation in adipose tissue overlapped directly or via proxy SNPs to previously reported disease loci of obesity related traits, including *CETP* and *FADS2*, which are both known to be associated with total cholesterol, LDL, HDL and triglyceride levels [47] These data support that genetic and epigenetic variation together influence metabolic phenotypes and disease risk in humans.

In order to provide further insights into mechanisms of genetic and epigenetic interaction and its impact on regulation of metabolic phenotypes, we used the CIT [21,70]. We discovered 39 significant mQTLs where DNA methylation represents the mediator between genetic loci and a metabolic trait. One of these mQTLs SNPs is associated with HDL regulation through DNA methylation of a CpG site annotated to *MCM6*. This is an MCM (minichromosome maintenance) complex gene that previously has been shown to affect total cholesterol levels [71]. Among other genes identified in the CIT analysis were *TCF19*, which has been associated with type 1 diabetes through GWAS [72], and *CAMK1D*, which has been associated with type 2 diabetes [73] This supports the role of DNA methylation as a direct mediator between genetic variation and metabolic phenotypes. However, while the majority of significant *cis*-mQTL SNP-CpG pairs were found to have independent effects on the analyzed phenotypes, the independence cannot be concluded due to some limitations in our analysis. First, only a few phenotypes were considered in the course of the analysis, and it might require other phenotypes to discover all cause–effect relationships between SNPs, methylation and metabolic phenotype. Second, as CIT only considers one SNP and one CpG site at a time, more complex interactions involving several SNPs and or CpGs can be missed, which suggests that more sophisticated analytical methods should be developed. An additional drawback of our study is the small sample size relative to the number of statistical comparisons. As the number of analyzed SNP–CpG pairs is in the order of 10^{11}, only the strongest interaction effects can be detected by means of our mQTL analysis. To reduce the number of type 2 errors during multiple testing procedures, we implemented a modified Bonferroni correction method, which took into account a linkage disequilibrium dependency between analyzed SNPs. Additionally, we performed eQTL and

CIT analysis only on the data that was shown to be significant in the mQTL analysis, thus again significantly reducing number of performed statistical tests. While eQTL analyses have been used to identify causal genetic variants for metabolic disease [74], here we provide the first CIT analysis of genetic variation, DNA methylation in adipose tissue and metabolic traits. Importantly, this analysis demonstrates how genetic variants mediate their effects on metabolic traits (e.g. BMI, cholesterol, HDL, HbA1c and HOMA-IR) via altered DNA methylation in human adipose tissue.

Interestingly, SNPs throughout the genome may introduce or delete CpG sites and thereby affect the possibility for DNA methylation to take place [22]. These so called CpG-SNPs are likely to show strong correlations with the degree of methylation in the SNP site. Indeed, here we found 447 CpG-SNPs associated with DNA methylation in adipose tissue.

Furthermore, we were able to replicate numerous of our unique CpG sites of significant *cis*-mQTLs in a study by Grundberg *et al.* [16] confirming the biological importance of our results. While both our study and Grundberg *et al.* performed mQTL analyses in human adipose tissue using the Illumina 450K array for DNA methylation and thereby comparable, divergence in *cis* boundary, sex and correction methods for multiple testing may explain some of the different results between the studies. It should also be noted that 39,386 of our significant *cis*-mQTLs in human adipose tissue were previously also identified in human pancreatic islets [20]. While this finding shows that some SNPs affect the DNA methylation pattern in multiple tissues, additional mQTL studies using the 450k array are needed in other tissues to test if the same associations are seen there.

We provide for the first time a combined genome-wide *cis*- and *trans*-mQTL analysis in human adipose tissue covering most genes and genomic regions. Our study demonstrates that interactions between genetic and epigenetic variation influences gene expression, molecular phenotypes and metabolic traits related to complex diseases in humans. We also provide details on potential causal relationships between genetic and epigenetic variation on metabolic phenotypes. Thus, DNA methylation variation may be of high importance in genetic association studies and may improve our understanding of molecular pathways in the context of complex human metabolic diseases.

Materials and Methods

Study samples and phenotypes

This study includes a total of 119 Scandinavian men without known disease. Their characteristics are presented in **Table 1**. The cohort includes subjects from four sub-cohorts, all previously described [15,48,75–77] and with DNA available from subcutaneous adipose tissue biopsies taken in the fasted state. The characteristics of the four sub-cohorts are presented separately in **S21 Table.** All study participants underwent a physical examination including measurements of BMI, waist and WHR. Moreover, blood sampling for analysis of lipids, glucose and insulin were done during the fasting state. Written informed consent was obtained from all participants and the research protocols were approved by the local human research ethics committees: Dnr 13/2006 (Lund University), Dnr 461/2006 (Lund University), KA 03129gm (Köpenhavns AMT). While three of sub-cohorts are intervention studies [75–77], one sub-cohort is a case-control cohort [15]. Only baseline samples from healthy subjects were included in this study.

Genotype data

Genotyping was performed in DNA extracted from blood of the 121 Scandinavian men using Illumina HumanOmniExpress BeadChip, which is a genome-wide array covering 731,412

SNPs, together with the iScan system (Illumina, San Diego, CA, USA). Genomic DNA was extracted from blood using the Gentra Puregene Blood Kit (Qiagen, Hilden, Germany). Genotypes were called using GenomeStudio® software (Illumina). All subjects passed call rate threshold of $> 98\%$. Sex discrepancy between reported sex and genotypic sex based on X-chromosome heterozygosity was detected for two subjects and these subjects were excluded from subsequent analyses. No subjects were found to be population outliers based on a population stratification test. SNPs were excluded if missing calls $> 5\%$, Hardy-Weinberg Equilibrium p-value < 0.001 and minor allele frequency < 0.05. Overall, 592,794 SNPs for 119 subjects passed quality control and were used for subsequent analyses. All genotype data were analyzed using Plink software (http://pngu.mgh.harvard.edu/purcell/plink/) [78].

DNA methylation data

Genome-wide DNA methylation profiling was performed in genomic DNA extracted using Qiagen DNA extraction kits (Qiagen) from adipose tissue from 119 Scandinavian men using the Infinium HumanMethylation450 BeadChip (Illumina). The DNA methylation array targets 485,577 probes across the genome, covering 99% of RefSeq genes and 96% of CpG islands. Genomic DNA (500 ng) from adipose tissue was bisulfite treated using the EZ DNA methylation kit (Zymo Research, Orange, CA, USA). DNA methylation analysis of bisulfite treated DNA was carried out with Infinium® assay following the standard Infinium HD Assay Methylation Protocol Guide (Part #15019519). BeadChips were scanned with Illumina iScan and raw data was imported to the GenomeStudio Methylation module software for calculation of methylation scores represented as methylation β-values. In sample quality control, all samples passed GenomeStudio quality control steps for bisulfite conversion efficiency, staining, hybridization, extension and specificity.

Individual probes with a mean Illumina detection p-value > 0.01 were considered not detected and subsequently excluded from further analysis. Non-CpG methylation probes and SNP-probes included on the array were also filtered out. After these quality control steps and after filtering DNA methylation data, 477,891 CpG sites remained for all included samples. Before further analysis, the DNA methylation data was exported from GenomeStudio and subsequently analyzed using Lumi package from Bioconductor [79]. Extracted methylation data were then converted from β-values to M-values [80], $M = log_2\left(\frac{\max(M,0)+1}{\max(U,0)+1}\right)$, where M and U are methylated and unmethylated channel intensities, respectively. The data was further background corrected and quantile normalized using lumi package [81]. To correct for batch effects, COMBAT normalization method [82] was used.

mRNA expression data

Genome-wide mRNA expression profiling using the whole-transcript GeneChip® Human Gene 1.0 ST Array (Affymetrix, Santa Clara, CA, USA) following the Affymetrix standard protocol was performed in RNA extracted from the subcutaneous adipose tissue biopsies of 118 out of 119 Scandinavian men using miRNeasy kit followed by the RNeasy MiniElute Cleanup Kit (Qiagen) or using the RNeasy Lipid Tissue Mini Kit (Qiagen). The array data was background corrected, quantile normalized and summarized with robust multichip average (RMA) procedure using oligo package [83] from Bioconductor. Normalized dataset was batch corrected using COMBAT [82]. In total, mRNA expression of 28,779 transcripts was obtained for subsequent analyses.

mQTL analysis

Associations between SNPs and DNA methylation of CpG sites were modeled as a linear relationship using DNA methylation levels as a dependent variable, SNP genotypes encoded as 0, 1 or 2 according to number of minor alleles. Due to the fact that both BMI and age can affect DNA methylation and, therefore, the association between SNP and DNA methylation, age, BMI and the sub-cohort were included as covariates. Calculations of associations were performed using the MatrixEQTL library for R programming language [84].

To distinguish between local (*cis*) and distant (*trans*) mQTLs a distance less or equal to 500 kb between a SNP and CpG site was used to define *cis*-mQTLs. All remaining SNP-CpG pairs were considered *trans*-mQTLs. In total we found 283,290,917,454 CpG–SNP pairs in the dataset, where 112,842,462 pairs were defined to be located in *cis* and 283,178,074,992 in *trans*. The *cis*- and *trans*-mQTL analyses were performed separately. In order to correct for multiple testing, p-value significance threshold was set, accounting for number of tests performed as well as the dependency of linkage disequilibrium (LD) between SNPs. LD-based SNP pruning was used to take into account the linkage dependency of SNPs that are run against the same quantitative trait locus in the mQTL analysis by calculating the number of independent tests based on $r^2<0.9$ for the SNPs. In the *cis*-analysis, LD based pruning of SNPs within a distance of 500 kb from a CpG site was performed by pairwise-tagging ($r^2<0.9$) and the total sum of all tag SNPs connected to each CpG site was used as correction value when correcting for multiple testing. LD calculations were performed using R trio package [85]. The correction value for the *trans*-analysis was calculated as the total number of analyzed CpG sites multiplied by the number of tag SNPs in the whole dataset (pairwise-tagging $r^2<0.9$) and subtracted by the correction value for the *cis*-analysis. Significance threshold was set to $p<0.05$ after correction for multiple testing. All SNPs connected to each CpG site after LD-based pruning were summed and the remaining number of 104,023,091 SNP-CpG pairs was used as correction value for multiple testing in *cis*. This resulted in a significance threshold of $0.05/104,023,091 = 4.8x10^{-10}$ in *cis*. In the *trans*-mQTL analysis, after LD-based pruning, 211,781,637,483 SNP-CpG pairs remain and this number was used as correction value for multiple testing. This resulted in a significance threshold of $0.05/211,781,637,483 = 2.3x10^{-13}$ in *trans*.

Impact of significant mQTL SNPs on mRNA expression

The relationship between SNPs found to be significantly associated with DNA methylation in the mQTL analysis and mRNA expression was tested in 118 of the men included in the study using a linear regression model with mRNA expression as a dependent variable, SNP genotypes encoded as 0, 1 or 2 according to number of minor alleles, and age, BMI and sub-cohort as covariates. Significant SNPs identified in the *cis*-mQTL analysis were only related to mRNA expression transcripts of genes located within 500 kb from respective SNP (*cis*). Significant SNPs identified in the *trans*-mQTL were related to mRNA expression transcripts of all analyzed genes. In total, 1,164,807 SNP-mRNA transcript combinations were found for significant *cis*-mQTLs, and 78,710,565 SNP-mRNA transcript combinations were found for significant *trans*-mQTLs. Correction value for multiple testing in the eQTL analysis was then calculated in similar way as for the mQTL analysis taking LD-based SNP pruning ($r^2<0.9$) into account. In the eQTL analysis of significant *cis*-mQTL SNPs, the number of LD pruned SNPs ($r^2<0.9$) to each mRNA transcript within 500 kb were summed up and used as the correction value for multiple testing. After LD-based pruning, 934,021 SNP-mRNA transcripts remain. This resulted in a significance threshold of $0.05/934,021 = 5.4x10^{-8}$ in *cis*. In the eQTL analysis of significant *trans*-mQTL SNPs, the correction value for multiple testing was calculated as the number of all *trans*-mQTL SNPs pruned for LD ($r^2<0.9$) multiplied by total number of

analyzed mRNA transcripts giving a remaining number of 33,326,082 SNP-mRNA transcripts. This resulted in a significance threshold of $0.05/33,326,082 = 1.5x10^{-9}$ in *trans*.

Impact of mQTL SNPs on metabolic phenotypes

The impact of identified SNPs in significant mQTLs on the following phenotypes; BMI, WHR, cholesterol, triglycerides, HDL, LDL, fasting glucose, fasting insulin, HOMA-B, HOMA-IR and HbA1c, was tested in 119 Scandinavian men included in this study. Associations between identified SNPs in the significant mQTLs and metabolic phenotypes were modeled as a linear relationship using metabolic phenotypes as the dependent variable, SNP genotypes encoded as 0, 1 or 2 according to number of minor alleles, and age and sub-cohort included as covariates in all the analyses. BMI was also included as a covariate when analyzing associations between SNPs and fasting glucose, fasting insulin, HOMA IR, HOMA-B and HbA1c. Traits for fasting insulin, HOMA-B and HOMA-IR have been naturally log transformed in the study cohort before analyses. Identified mQTL SNPs showing association to a metabolic phenotype in our study cohort ($p<0.05$), were also looked-up in public available GWAS data from the GIANT consortium [43,44], MAGIC investigators [36,45,46] and GLGC consortium [47], for respective trait. SNPs showing association to a metabolic phenotype with the same allelic effect sign and with p-value <0.05 in both our study cohort and consortia data were considered detected.

Overlap between mQTL SNPs and public available GWAS data

The catalog of published GWAS data was used to search for SNPs reported to be associated with obesity, type 2 diabetes and related metabolic traits ($p<10^{-5}$). To increase reference coverage for overlap between datasets of identified mQTL SNPs and identified SNPs reported in GWAS catalog, a SNP annotation and proxy (SNAP) search [86] was performed to identify SNPs in LD with the identified mQTL SNPs. The proxy search was based on pairwise LD calculations of genotype data from the 1000 Genomes project of the CEU population panel with $r^2>0.8$ and distance limit of 500 kb from the query SNP.

Causal Inference Test (CIT)

The CIT was used to test if DNA methylation is a mediator between genotype variation and a phenotypic trait [21]. The causality can be inferred if all of the following are true: 1) G and M are associated, 2) G and P are associated, 3) G is associated with M|P and 4) G is independent of P|M, where G is a genotype marker, M is a DNA methylation measure and P is a phenotypic trait, provided that G is randomized [21]. Causal role of DNA methylation is inferred if p-value for causal relationship hypothesis is less than 0.05.

Statistical analysis

Data were analyzed using linear regression models, Pearson chi-squared test or Fisher's exact test. All statistical calculations were performed using R programming language [87]. Results are expressed as Box and Whiskers plots. Pathway analysis using WebGestalt [26].

Supporting Information

S1 Table. Identified *cis*-mQTLs. Sheet a: Identified *cis*-mQTL SNP-CpG pairs, including chromosomal location and relation to CpG islands and gene regions. Sheet b: SNP-CpG pairs where SNP is located in either C or G of the CpG site, so called CpG-SNPs. Sheet c: Additional annotation data for SNPs present in sheet a, based on HumanOmniExpress-12v1_J_Gene_Annotation_build37 (Illumina). Sheet d: Additional annotation data for CpGs present in sheet a,

based on Infinium HumanMethylation 450 BeadChip [25] and probe cross-reactivity info as reported by Chen et al [88].
(XLSX)

S2 Table. Identified *trans*-mQTLs. Sheet a: Identified *trans*-mQTL SNP-CpG pairs, including statistical results of associations and chromosomal location and relation to CpG islands and gene regions. Sheet b: Additional annotation data for SNPs present in sheet a, based on HumanOmniExpress-12v1_J_Gene_Annotation_build37 (Illumina). Sheet c: Additional annotation data for CpGs present in sheet a, based on Infinium HumanMethylation 450 BeadChip [25] and probe cross-reactivity info as reported by Chen et al [88]
(XLSX)

S3 Table. KEGG pathways identified among genes annotated to significant *cis* mQTL CpG sites. Sheet a: **KEGG** pathways enriched among genes annotated to CpG sites from significant *cis*-mQTLs. Sheet b: **KEGG** pathways enriched among genes annotated to CpG sites from significant *trans*-mQTLs.
(XLSX)

S4 Table. Identified *cis*-mQTL SNPs that are also reported as disease SNPs in GWAS catalog [27]. SNPs that are found directly in the catalog are marked with grey, and ones that are found to be in LD with GWAS catalog SNPs with white. LD proxy analysis performed using SNAP (1000 Genomes project, CEU population panel, r2 > 0.8, distance limit 500kb) [86].
(XLSX)

S5 Table. Identified *trans*-mQTL SNPs that are also reported as disease SNPs in GWAS catalog [27]. SNPs that are found directly in the catalog are marked with grey, and ones that are found to be in LD with GWAS catalog SNPs with white. LD proxy analysis performed using SNAP (1000 Genomes project, CEU population panel, r2 > 0.8, distance limit 500kb) [86].
(XLSX)

S6 Table. Identified significant *cis*-mQTL SNPs that also show associations with gene expression. Sheet a: Identified *cis* SNP-mRNA transcript pairs, including statistical results of associations and gene assignment for mRNA transcripts. Only pairs with p-value < 0.05 after multiple testing corrections are included. Sheet b: Additional annotation data for SNPs present in sheet a, based on HumanOmniExpress-12v1_J_Gene_Annotation_build37 (Illumina). Sheet c: Additional annotation data for probesets present in sheet a. Annotations are based on NetAffx transcript cluster data for HuGene-1_0-st array (Affymetrix).
(XLSX)

S7 Table. Identified significant *trans*-mQTL SNPs that also show associations with gene expression. Sheet a: Identified *trans* SNP-mRNA transcript pairs, including statistical results of associations and gene assignment for mRNA transcripts. Sheet b: Additional annotation data for SNPs present in sheet a. Sheet c: Additional annotation data for probesets present in sheet a. Annotations are based on NetAffx transcript cluster data for HuGene-1_0-st array (Affymetrix).
(XLSX)

S8 Table. Identified *cis*-mQTL SNPs that show association with metabolic phenotypes in the study cohort ($p<0.05$) and are also identified in MAGIC, GIANT, or GLGC consortia ($p<0.05$). Excel table representing *cis*-mQTL SNPs that are identified in MAGIC, GIANT, or GLGC consortia ($p<0.05$) overlapping with SNPs that show association with metabolic phenotypes in the study cohort ($p<0.05$). Sheet a: SNPs associated with BMI in study cohort and

GIANT consortium [44]. Sheet b: SNPs associated with Waist-hip ratio in study cohort and GIANT consortium [43]. Sheet c: SNPs associated with Fasting glucose in study cohort and MAGIC consortium [45]. Sheet d: SNPs associated with Fasting insulin in study cohort and MAGIC consortium [45]. Sheet e: SNPs associated with HOMA-B in study cohort and MAGIC consortium [45]. Sheet f: SNPs associated with HOMA-IR in study cohort and MAGIC consortium [45]. Sheet g: SNPs associated with HbA1c in study cohort and MAGIC consortium [46]. Sheet h: SNPs associated with Total cholesterol in study cohort and GLGC consortium [47]. Sheet i SNPs associated with Triglycerides in study cohort and GLGC consortium [47]. Sheet j: SNPs associated with HDL in study cohort and GLGC consortium [47]. Sheet k: SNPs associated with LDL in study cohort and consortium GLGC [47].
(XLSX)

S9 Table. Significant *cis*-mQTL/eQTL SNPs that show association with metabolic phenotypes in the study cohort and are also identified in MAGIC, GIANT, or GLGC consortia ($p<0.05$).
(XLSX)

S10 Table. Identified *trans*-mQTL SNPs that show association with metabolic phenotypes in the study cohort ($p<0.05$) and are also identified in MAGIC, GIANT, or GLGCR consortia ($p<0.05$). Sheet a: SNPs associated with BMI in study cohort and GIANT consortium [44]. Sheet b: SNPs associated with Waist-hip ratio in study cohort and GIANT consortium [43]. Sheet c: SNPs associated with Fasting glucose in study cohort and MAGIC consortium [45]. Sheet d: SNPs associated with Fasting insulin in study cohort and MAGIC consortium [45]. Sheet e: SNPs associated with HOMA-B in study cohort and MAGIC consortium [45]. Sheet f: SNPs associated with HOMA-IR in study cohort and MAGIC consortium [45]. Sheet g: SNPs associated with HbA1c in study cohort and MAGIC consortium [46]. Sheet h: SNPs associated with Total cholesterol in study cohort and GLGC consortium [47]. Sheet i SNPs associated with Triglycerides in study cohort and GLGC consortium [47]. Sheet j: SNPs associated with HDL in study cohort and GLGC consortium [47]. Sheet k: SNPs associated with LDL in study cohort and consortium GLGC [47].
(XLSX)

S11 Table. Significant *cis*-mQTL SNPs associated with CRP in Denghan et al. [65].
(XLSX)

S12 Table. Significant *cis*-mQTL SNP-CpG pairs that are also reported to show significant associations in Olsson et al. [20].
(XLSX)

S13 Table. Significant *trans*-mQTL SNP-CpG pairs that are also reported to show significant associations in Olsson et al. [20].
(XLSX)

S14 Table. Significant *cis*- and *trans*- mQTL CpG sites that are also reported to show differential methylation in Nilsson et al. [15]. Sheet a: Identified *cis*-mQTL CpG sites that are reported in Nilsson et al. [15]. Sheet b: Identified *trans*-mQTL CpG sites that are reported in Nilsson et al. [15]. Sheet c: Additional annotation data for CpGs present in sheet a, based on Infinium HumanMethylation 450 BeadChip. [25]. Sheet d: Additional annotation data for SNPs present in sheet a, based on HumanOmniExpress-12v1_J_Gene_Annotation_build37 (Illumina).
(XLSX)

S15 Table. Significant *cis*- and *trans*- mQTL CpG sites that are also reported to show significant associations with BMI and HbA1c in Rönn et al. [48].
(XLSX)

S16 Table. *cis*-mQTLs identified in the analysis of subcohorts 1 and 2. Sheet a: Identified *cis*-mQTL SNP-CpG pairs, including chromosomal location and relation to CpG islands and gene regions. Sheet b: SNP-CpG pairs where SNP is located in either C or G of the CpG site, so called CpG-SNPs. Sheet c: Additional annotation data for SNPs present in sheet a, based on HumanOmniExpress-12v1_J_Gene_Annotation_build37 (Illumina). Sheet d: Additional annotation data for CpGs present in sheet a, based on Infinium HumanMethylation 450 BeadChip [25] and probe cross-reactivity info as reported by Chen et al [88].
(XLSX)

S17 Table. *trans*-mQTLs *cis*-mQTLs identified in the analysis of subcohorts 1 and 2. Sheet a: Identified *trans*-mQTL SNP-CpG pairs, including statistical results of associations and chromosomal location and relation to CpG islands and gene regions. Sheet b: Additional annotation data for SNPs present in sheet a, based on HumanOmniExpress-12v1_J_Gene_Annotation_build37 (Illumina). Sheet c: Additional annotation data for CpGs present in sheet a, based on Infinium HumanMethylation 450 BeadChip [25] and probe cross-reactivity info as reported by Chen et al [88].
(XLSX)

S18 Table. *cis*-mQTLs identified without BMI as a covariate. Sheet a: Identified *cis*-mQTL SNP-CpG pairs, including chromosomal location and relation to CpG islands and gene regions. Sheet b: SNP-CpG pairs where SNP is located in either C or G of the CpG site, so called CpG-SNPs. Sheet c: Additional annotation data for SNPs present in sheet a, based on HumanOmniExpress-12v1_J_Gene_Annotation_build37 (Illumina). Sheet d: Additional annotation data for CpGs present in sheet a, based on Infinium HumanMethylation 450 BeadChip [25] and probe cross-reactivity info as reported by Chen et al [88].
(XLSX)

S19 Table. *trans*-mQTLs identified without BMI as a covariate. Sheet a: Identified *trans*-mQTL SNP-CpG pairs, including statistical results of associations and chromosomal location and relation to CpG islands and gene regions. Sheet b: Additional annotation data for SNPs present in sheet a, based on HumanOmniExpress-12v1_J_Gene_Annotation_build37 (Illumina). Sheet c: Additional annotation data for CpGs present in sheet a, based on Infinium HumanMethylation 450 BeadChip [25] and probe cross-reactivity info as reported by Chen et al [88].
(XLSX)

S20 Table. Association between DNA methylation and gene expression in human adipose tissue.
(XLSX)

S21 Table. Sample characteristics of 4 different cohorts included in the study.
(XLSX)

Acknowledgments

This study wouldn't be possible without the help from participating clinicians, who performed clinical studies and provided the clinical materials, including Targ Elgzyri, Ylva Wessman, Marianne Modest and Lars Sander Koch. We thank SCIBLU (Swegene Center for Integrative

Biology at Lund University) Genomics Facility for help with DNA methylation and mRNA expression analyses.

This work was supported by grants from the Swedish Research Council, Region Skåne (ALF), Knut and Alice Wallenberg Foundation, Novo Nordisk Foundation, EFSD/Lilly Fellowship, Söderberg Foundation, The Swedish Diabetes foundation, Påhlsson Foundation, EXODIAB, Linné grant (B31 5631/2006), The Danish Strategic Research Council, The Danish Council for Independent Research, Rigshospitalet, University of Copenhagen, Steno Diabetes Center, Danish Diabetes Academy.

Data on glycaemic traits have been contributed by MAGIC investigators and have been downloaded from www.magicinvestigators.org

Author Contributions

Conceived and designed the experiments: PV AHO CL. Performed the experiments: L. Gillberg EN TR. Analyzed the data: PV AHO CL. Contributed reagents/materials/analysis tools: SWJ CB PAJ K-FE L. Groop AV. Wrote the paper: PV AHO CL.

References

1. McCarthy MI. Genomics, Type 2 Diabetes, and Obesity. N Engl J Med. 2010; 363: 2339–2350. doi: 10.1056/NEJMra0906948 PMID: 21142536
2. Franks PW, Ling C. Epigenetics and obesity: the devil is in the details. BMC Med. 2010; 8: 88. doi: 10.1186/1741-7015-8-88 PMID: 21176136
3. Ling C, Groop L. Epigenetics: A Molecular Link Between Environmental Factors and Type 2. Diabetes. 2009; 58: 2718–2725. doi: 10.2337/db09-1003 PMID: 19940235
4. Bird A. Perceptions of epigenetics. Nature. 2007; 447: 396–398. doi: 10.1038/nature05913 PMID: 17522671
5. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002; 16: 6–21. doi: 10.1101/gad.947102 PMID: 11782440
6. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012; 13: 484–492. doi: 10.1038/nrg3230 PMID: 22641018
7. Brøns C, Jacobsen S, Nilsson E, Rönn T, Jensen CB, Storgaard H, et al. Deoxyribonucleic acid methylation and gene expression of PPARGC1A in human muscle is influenced by high-fat overfeeding in a birth-weight-dependent manner. J Clin Endocrinol Metab. 2010; 95: 3048–3056. doi: 10.1210/jc.2009-2413 PMID: 20410232
8. Jacobsen SC, Brøns C, Bork-Jensen J, Ribel-Madsen R, Yang B, Lara E, et al. Effects of short-term high-fat overfeeding on genome-wide DNA methylation in the skeletal muscle of healthy young men. Diabetologia. 2012; 55: 3341–3349. doi: 10.1007/s00125-012-2717-8 PMID: 22961225
9. Nitert MD, Dayeh T, Volkov P, Elgzyri T, Hall E, Nilsson E, et al. Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. Diabetes. 2012; 61: 3322–3332. doi: 10.2337/db11-1653 PMID: 23028138
10. Rönn T, Volkov P, Davegårdh C, Dayeh T, Hall E, Olsson AH, et al. A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS Genet. 2013; 9: e1003572. doi: 10.1371/journal.pgen.1003572 PMID: 23825961
11. Anway MD, Cupp AS, Uzumcu M, Skinner MK. Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science. 2005; 308: 1466–1469. doi: 10.1126/science.1108190 PMID: 15933200
12. Chong S, Whitelaw E. Epigenetic germline inheritance. Curr Opin Genet Dev. 2004; 14: 692–696. doi: 10.1016/j.gde.2004.09.001 PMID: 15531166
13. Kaminsky ZA, Tang T, Wang S-C, Ptak C, Oh GHT, Wong AHC, et al. DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet. 2009; 41: 240–245. doi: 10.1038/ng.286 PMID: 19151718
14. Ollikainen M, Smith KR, Joo EJ-H, Ng HK, Andronikos R, Novakovic B, et al. DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Hum Mol Genet. 2010; 19: 4176–4188. doi: 10.1093/hmg/ddq336 PMID: 20699328

15. Nilsson E, Jansson PA, Perfilyev A, Volkov P, Pedersen M, Svensson MK, et al. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. Diabetes. 2014; doi: 10.2337/db13-1459

16. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet. 2013; 93: 876–890. doi: 10.1016/j.ajhg.2013.10.004 PMID: 24183450

17. Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, et al. Genetic control of individual differences in gene-specific methylation in human brain. Am J Hum Genet. 2010; 86: 411–419. doi: 10.1016/j.ajhg.2010.02.005 PMID: 20215007

18. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. 2011; 12: R10. doi: 10.1186/gb-2011-12-1-r10 PMID: 21251332

19. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai S- L, et al. Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain. PLoS Genet. 2010; 6: e1000952. doi: 10.1371/journal.pgen.1000952 PMID: 20485568

20. Olsson AH, Volkov P, Bacos K, Dayeh T, Hall E, Nilsson EA, et al. Genome-Wide Associations between Genetic and Epigenetic Variation Influence mRNA Expression and Insulin Secretion in Human Pancreatic Islets. PLoS Genet. 2014; 10: e1004735. doi: 10.1371/journal.pgen.1004735 PMID: 25375650

21. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. BMC Genet. 2009; 10: 23. doi: 10.1186/1471-2156-10-23 PMID: 19473544

22. Dayeh TA, Olsson AH, Volkov P, Almgren P, Rönn T, Ling C. Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia. 2013; 56: 1036–1046. doi: 10.1007/s00125-012-2815-7 PMID: 23462794

23. Pociot F, McDermott MF. Genetics of type 1 diabetes mellitus. Genes Immun. 2002; 3: 235–249. doi: 10.1038/sj.gene.6363875 PMID: 12140742

24. Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens. 2004; 64: 631–649. doi: 10.1111/j.1399-0039.2004.00327.x PMID: 15546336

25. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011; 98: 288–295. doi: 10.1016/j.ygeno.2011.07.007 PMID: 21839163

26. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res. 2013; 41: W77–83. doi: 10.1093/nar/gkt439 PMID: 23703215

27. Hindorff, L.A., MacArthur, J., Morales, J., Junkins H.A., Hall, P.N., Klemm, A.K., & Manolio, T.A. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed September 23 2013.

28. Farooqi S, O'Rahilly S. Genetics of obesity in humans. Endocr Rev. 2006; 27: 710–718. doi: 10.1210/er.2006-0040 PMID: 17122358

29. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010; 42: 937–948. doi: 10.1038/ng.686 PMID: 20935630

30. Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PloS One. 2012; 7: e51954. doi: 10.1371/journal.pone.0051954 PMID: 23251661

31. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PIW, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007; 316: 1331–1336. doi: 10.1126/science.1142358 PMID: 17463246

32. Lowe JK, Maller JB, Pe'er I, Neale BM, Salit J, Kenny EE, et al. Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae. PLoS Genet. 2009; 5: e1000365. doi: 10.1371/journal.pgen.1000365 PMID: 19197348

33. Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. Nat Genet. 2008; 40: 716–718. doi: 10.1038/ng.156 PMID: 18454146

34. Sabatti C, Service SK, Hartikainen A-L, Pouta A, Ripatti S, Brodsky J, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet. 2009; 41: 35–46. doi: 10.1038/ng.271 PMID: 19060910

35. Ryckman KK, Smith CJ, Jelliffe-Pawlowski LL, Momany AM, Berberich SL, Murray JC. Metabolic heritability at birth: implications for chronic disease research. Hum Genet. 2014; 133: 1049–1057. doi: 10.1007/s00439-014-1450-4 PMID: 24850141

36. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012; 44: 659–669. doi: 10.1038/ng.2274 PMID: 22581228

37. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet TIG. 2008; 24: 408–415. doi: 10.1016/j.tig.2008.06.001 PMID: 18597885

38. Zhu AZX, Renner CC, Hatsukami DK, Benowitz NL, Tyndale RF. CHRNA5-A3-B4 genetic variants alter nicotine intake and interact with tobacco use to influence body weight in Alaska Native tobacco users. Addict Abingdon Engl. 2013; 108: 1818–1828. doi: 10.1111/add.12250

39. Mahajan A, Sim X, Ng HJ, Manning A, Rivas MA, Highland HM, et al. Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. PLoS Genet. 2015; 11: e1004876. doi: 10.1371/journal.pgen.1004876 PMID: 25625282

40. Williams MJ, Almén MS, Fredriksson R, Schiöth HB. What model organisms and interactomics can reveal about the genetics of human obesity. Cell Mol Life Sci CMLS. 2012; 69: 3819–3834. doi: 10.1007/s00018-012-1022-5 PMID: 22618246

41. Fox CS, Liu Y, White CC, Feitosa M, Smith AV, Heard-Costa N, et al. Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. PLoS Genet. 2012; 8: e1002695. doi: 10.1371/journal.pgen.1002695 PMID: 22589738

42. Mackay DJG, Callaway JLA, Marks SM, White HE, Acerini CL, Boonen SE, et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet. 2008; 40: 949–951. doi: 10.1038/ng.187 PMID: 18622393

43. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet. 2010; 42: 949–960. doi: 10.1038/ng.685 PMID: 20935629

44. Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI, et al. FTO genotype is associated with phenotypic variability of body mass index. Nature. 2012; 490: 267–272. doi: 10.1038/nature11401 PMID: 22982992

45. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010; 42: 105–116. doi: 10.1038/ng.520 PMID: 20081858

46. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al. Common Variants at 10 Genomic Loci Influence Hemoglobin A1C Levels via Glycemic and Nonglycemic Pathways. Diabetes. 2010; 59: 3229–3239. doi: 10.2337/db10-0502 PMID: 20858683

47. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013; 45: 1274–1283. doi: 10.1038/ng.2797 PMID: 24097068

48. Rönn T, Volkov P, Gillberg L, Kokosar M, Perfilyev A, Jacobsen AL, et al. Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Hum Mol Genet. 2015; 24: 3792–3813. doi: 10.1093/hmg/ddv124 PMID: 25861810

49. Groop L, Pociot F. Genetics of diabetes—are we missing the genes or the disease? Mol Cell Endocrinol. 2014; 382: 726–739. doi: 10.1016/j.mce.2013.04.002 PMID: 23587769

50. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet. 2007; 39: 770–775. doi: 10.1038/ng2043 PMID: 17460697

51. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007; 316: 889–894. doi: 10.1126/science.1141634 PMID: 17434869

52. Kilpeläinen TO, Zillikens MC, Stančáková A, Finucane FM, Ried JS, Langenberg C, et al. Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. Nat Genet. 2011; 43: 753–760. doi: 10.1038/ng.866 PMID: 21706003

53. Gluckman PD, Hanson MA, Buklijas T, Low FM, Beedle AS. Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. Nat Rev Endocrinol. 2009; 5: 401–408. doi: 10.1038/nrendo.2009.102 PMID: 19488075

54. Ling C, Del Guerra S, Lupi R, Rönn T, Granhall C, Luthman H, et al. Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia. 2008; 51: 615–622. doi: 10.1007/s00125-007-0916-5 PMID: 18270681

55. Yang BT, Dayeh TA, Kirkpatrick CL, Taneera J, Kumar R, Groop L, et al. Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. Diabetologia. 2011; 54: 360–367. doi: 10.1007/s00125-010-1967-6 PMID: 21104225

56. Yang BT, Dayeh TA, Volkov PA, Kirkpatrick CL, Malmgren S, Jing X, et al. Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. Mol Endocrinol Baltim Md. 2012; 26: 1203–1212. doi: 10.1210/me.2012-1004

57. Barrès R, Osler ME, Yan J, Rune A, Fritz T, Caidahl K, et al. Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. Cell Metab. 2009; 10: 189–198. doi: 10.1016/j.cmet.2009.07.011 PMID: 19723495

58. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aïssi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. Lancet. 2014; 383: 1990–1998. doi: 10.1016/S0140-6736(13)62674-4 PMID: 24630777

59. Dayeh T, Volkov P, Salö S, Hall E, Nilsson E, Olsson AH, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. PLoS Genet. 2014; 10: e1004160. doi: 10.1371/journal.pgen.1004160 PMID: 24603685

60. Ling C, Poulsen P, Simonsson S, Rönn T, Holmkvist J, Almgren P, et al. Genetic and epigenetic factors are associated with expression of respiratory chain component NDUFB6 in human skeletal muscle. J Clin Invest. 2007; 117: 3427–3435. doi: 10.1172/JCI30938 PMID: 17948130

61. Drong AW, Nicholson G, Hedman ÅK, Meduri E, Grundberg E, Small KS, et al. The Presence of Methylation Quantitative Trait Loci Indicates a Direct Genetic Influence on the Level of DNA Methylation in Adipose Tissue. PLoS ONE. 2013; 8: e55923. doi: 10.1371/journal.pone.0055923 PMID: 23431366

62. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 2006; 38: 1378–1385. doi: 10.1038/ng1909 PMID: 17072317

63. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009; 41: 178–186. doi: 10.1038/ng.298 PMID: 19151715

64. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;advance online publication. doi: 10.1038/nature12433

65. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation. 2011; 123: 731–738. doi: 10.1161/CIRCULATIONAHA.110.948570 PMID: 21300955

66. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010; 42: 579–589. doi: 10.1038/ng.609 PMID: 20581827

67. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010; 42: 937–948. doi: 10.1038/ng.686 PMID: 20935630

68. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012; 44: 991–1005. doi: 10.1038/ng.2385 PMID: 22885924

69. Feinberg AP, Irizarry RA. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci. 2009; 200906183. doi: 10.1073/pnas.0906183107

70. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol. 2013; 31: 142–147. doi: 10.1038/nbt.2487 PMID: 23334450

71. Ma L, Yang J, Runesha HB, Tanaka T, Ferrucci L, Bandinelli S, et al. Genome-wide association analysis of total cholesterol and high-density lipoprotein cholesterol levels using the Framingham Heart Study data. BMC Med Genet. 2010; 11: 55. doi: 10.1186/1471-2350-11-55 PMID: 20370913

72. Cheung YH, Watkinson J, Anastassiou D. Conditional meta-analysis stratifying on detailed HLA genotypes identifies a novel type 1 diabetes locus around TCF19 in the MHC. Hum Genet. 2011; 129: 161–176. doi: 10.1007/s00439-010-0908-2 PMID: 21076979

73. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008; 40: 638–645. doi: 10.1038/ng.120 PMID: 18372903

74. Das SK, Sharma NK. Expression quantitative trait analyses to identify causal genetic variants for type 2 diabetes susceptibility. World J Diabetes. 2014; 5: 97–114. doi: 10.4239/wjd.v5.i2.97 PMID: 24748924

75. Elgzyri T, Parikh H, Zhou Y, Dekker Nitert M, Rönn T, Segerström ÅB, et al. First-degree relatives of type 2 diabetic patients have reduced expression of genes involved in fatty acid metabolism in skeletal muscle. J Clin Endocrinol Metab. 2012; 97: E1332–1337. doi: 10.1210/jc.2011-3037 PMID: 22547424

76. Brøns C, Jensen CB, Storgaard H, Alibegovic A, Jacobsen S, Nilsson E, et al. Mitochondrial function in skeletal muscle is normal and unrelated to insulin action in young men born with low birth weight. J Clin Endocrinol Metab. 2008; 93: 3885–3892. doi: 10.1210/jc.2008-0630 PMID: 18628517

77. Jørgensen SW, Brøns C, Bluck L, Hjort L, Færch K, Thankamony A, et al. Metabolic response to 36 hours of fasting in young men born small vs appropriate for gestational age. Diabetologia. 2014; doi: 10.1007/s00125-014-3406-6

78. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81: 559–575. doi: 10.1086/519795 PMID: 17701901

79. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5: R80. doi: 10.1186/gb-2004-5-10-r80 PMID: 15461798

80. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010; 11: 587. doi: 10.1186/1471-2105-11-587 PMID: 21118553

81. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinforma Oxf Engl. 2008; 24: 1547–1548. doi: 10.1093/bioinformatics/btn224

82. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007; 8: 118–127. doi: 10.1093/biostatistics/kxj037 PMID: 16632515

83. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinforma Oxf Engl. 2010; 26: 2363–2367. doi: 10.1093/bioinformatics/btq431

84. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012; 28: 1353–1358. doi: 10.1093/bioinformatics/bts163 PMID: 22492648

85. Holger Schwender, Qing Li, Christoph Neumann, Margaret Taub, Ingo Ruczinski. trio: Testing of SNPs and SNP Interactions in Case-Parent Trio Studies. R package version 3.0.0. 2013;

86. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, Bakker PIW de. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008; 24: 2938–2939. doi: 10.1093/bioinformatics/btn564 PMID: 18974171

87. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013;

88. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013; 8: 203–209. doi: 10.4161/epi.23470 PMID: 23314698

Study IV

Whole-genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis

Petr Volkov[1], Karl Bacos[1], Jones K. Ofori[2], Jonathan Lou S. Esguerra[2], Lena Eliasson[2], Tina Rönn[1*] and Charlotte Ling[1*]

[1]Epigenetics & Diabetes unit, Department of Clinical Sciences, Lund University Diabetes Centre, Scania University Hospital, Malmö, Sweden.

[2]Islet cell exocytosis unit, Department of Clinical Sciences, Lund University Diabetes Centre, Scania University Hospital, Malmö, Sweden.

* Equal contribution

Corresponding author:

Charlotte Ling

e-mail: charlotte.ling@med.lu.se

Phone nr: +46 40 391213

Short title: WGBS of human islets

Abstract

Background: Current knowledge about the role of epigenetics in type 2 diabetes (T2D) remains limited. Only a few studies have investigated DNA methylation of selected candidate genes or a very small fraction of genomic CpG sites in human pancreatic islets, the tissue of primary pathogenic importance for diabetes. Our aim was to characterize the whole-genome DNA methylation landscape in human pancreatic islets, to identify differentially methylated regions (DMRs) in diabetic islets, and to investigate the function of DMRs in islet biology.

Methods and findings: Here, we performed whole-genome bisulfite sequencing, which is a comprehensive and unbiased method to study DNA methylation throughout the genome on a single nucleotide resolution, in pancreatic islets from donors with T2D and non-diabetic controls. We identified 25,820 DMRs in islets from individuals with T2D. These novel DMRs cover loci with known islet function e.g. *PDX1*, *TCF7L2* and *ADCY5*. Importantly, binding sites for islet specific transcription factors, enhancer regions and different histone marks were enriched in the T2D associated DMRs. We also identified 457 genes, including *NR4A3*, *PARK2*, *PID1*, *SLC2A2* and *SOCS2* that had both DMRs and significant expression changes in T2D islets. To mimic the situation in T2D, candidate genes were overexpressed or silenced in β-cells. This resulted in impaired insulin secretion, thereby connecting differential methylation to islet dysfunction. We further explored the islet methylome and found a strong link between methylation levels and histone marks. Additionally, DNA methylation in different genomic regions and of different transcript types (i.e. protein-coding, non-coding and pseudogenes) associated with islet expression levels.

Conclusions: Our study provides a comprehensive picture of the islet DNA methylome in both non-diabetic and diabetic individuals and highlights the importance of epigenetic dysregulation in pancreatic islets and T2D pathogenesis.

Introduction

Impaired insulin secretion is a key feature of type 2 diabetes (T2D). However, the molecular mechanisms underlying pancreatic islet dysfunction in patients with T2D are largely unknown. While genetic risk factors are known to contribute to the etiology of T2D, less than 20% of the estimated heritability of the disease can be explained by single nucleotide polymorphisms (SNPs) identified by genome-wide association studies (GWAS) [1]. As a consequence, one must look elsewhere to find disease-causing mechanisms. Given the important role of environmental factors in the pathogenesis of T2D, mechanisms mediating the interaction between the environment and the genome, such as epigenetic mechanisms, may be of particular importance. Studying epigenetic processes in the tissue of primary pathogenetic importance, the pancreatic islets, may reveal mechanisms of significance for T2D. Indeed, our group and others have identified altered DNA methylation patterns in pancreatic islets from subjects with T2D compared with non-diabetic controls [2-6]. However, the method used in our previous studies, Infinium HumanMethylation450K BeadChip, only covers ~1.5% of the CpG sites in the human genome [7]. Therefore, to obtain a more complete picture of the methylome in human pancreatic islets and to further dissect the impact of epigenetics in T2D, genome-wide analyses covering the majority of methylation sites are needed.

Whole-genome bisulfite sequencing (WGBS) is the most comprehensive method to study DNA methylation throughout the genome on a single nucleotide resolution. This method is costly and most studies have so far been based on very few samples in a selected set of human tissues [8-17]. Despite the limited data sets, they have provided knowledge regarding gene regulation and the distribution of DNA methylation across different genomic regions and cell types. The largest effort to describe the human epigenome has been undertaken by the NIH Roadmap Epigenomics Consortium, including WGBS of 37 human epigenomes in various tissues and cell types [18]. However, previous WGBS studies did not include human pancreatic islets, nor did they examine the role of DNA methylation in a T2D case-control cohort.

To address this knowledge gap and dissect epigenetic alterations in T2D, we performed WGBS of DNA from human pancreatic islets obtained from 14 donors, eight normoglycemic controls and six diagnosed

with T2D. We identified numerous novel differentially methylated regions (DMRs) in pancreatic islets from donors with T2D, including several loci known to be important for islet function. We further characterized the methylome in human pancreatic islets and studied the relation between DNA methylation and histone modifications, enhancer regions, RNA expression and transcription factor binding sites.

Methods and Materials

Human pancreatic islet samples

Human pancreatic islets were obtained from the Nordic Network for Islet Transplantation, Uppsala University, Sweden. The pancreatic islet donor or her/his relatives had upon admission to intensive care unit given their informed consent to donate organs for medical research. All procedures were approved by ethics committees at Uppsala and Lund Universities.

Pancreatic islets were prepared and cultured as previously described [19] followed by RNA and DNA isolation using the AllPrep DNA/RNA Mini Kit (Qiagen GmbH, Hilden, Germany). Concentrations and purity of RNA and DNA were measured using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

Pancreatic islets of eight normoglycemic control donors and six donors diagnosed with T2D were used for WGBS (**Table 1**). These islets were cultured for 3.1 ± 0.3 days and the islet purity was 82.8 ± 2.2% based on dithizone staining. Infinium HumanMethylation450K BeadChip (Illumina, San Diego, CA, USA) DNA methylation data and RNA sequencing data were also generated for these 14 islet donors.

Table 1. Characteristics for donors of pancreatic islets included in the WGBS analysis.

	Controls (*n*=8)	Type 2 diabetic (*n*=6)	*P*-value
Sex (m/f)	4/4	3/3	
Age (years)	52.5 ± 3.2 (40-67)	58.2 ± 3.6 (45-66)	0.26
BMI (kg/m^2)	24.9 ± 0.3 (23.9-26.6)	28.0 ± 2.0 (22.9-34.6)	0.10
HbA1c (%)	5.47 ± 0.10 (*n*=7; 5.2-6.0)	7.12 ± 0.21 (6.3-7.8)	< 0.0001

Data are presented as mean ± SEM (range). t-tests and two-tailed *P*-values were used to detect differences between groups.

In the present study, we also examined Infinium 450K DNA methylation data from our previous study including islets from 15 type 2 diabetic and 34 control donors [2] as well as RNA sequencing data from

89 islet donors previously described [20]. Biological replication was performed on pancreatic islets from 19 donors diagnosed with T2D and 56 normoglycemic controls not included in the WGBS analysis (**Table S1**).

Library preparation and WGBS

300 ng DNA isolated from pancreatic islets was bisulfite treated with the EZ DNA Methylation Gold D5005 kit (Zymo Research Corporation, Irvine, CA, USA) according to the manufacturers' protocol. For each of the islet samples, triplicate sequencing libraries were prepared using the TruSeq (EpiGnome) DNA Methylation Kit (Illumina, EGMK91324) according to the manufacturers' protocol (#15066014 revA). 100 ng bisulfite treated islet DNA was used to prepare each TruSeq sequencing library. Next, these triplicate libraries for each sample were pooled and sequenced using Illumina HiSeq2500 utilizing 125 bp long paired end reads of Illumina type 4 chemistry. For one of the samples, a technical replication was performed by sequencing in total six lanes for three libraries prepared using NEXTflex™ Bisulfite Library Prep Kit (Bioo Scientific, Austin, TX, USA). Here, 1000 ng bisulfite treated islet DNA was used to prepare each NEXTflex™ sequencing library.

WGBS data analysis and DMR calling

WGBS data in FASTQ format generated using the Illumina HiSeq platform were used for further analyses. Illumina adapter sequences were trimmed from 3' and 5' ends of all paired-end reads, and bases with quality Phred score less than 20 were filtered using Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Next, reads shorter than 20 bp after trimming were filtered out, again using Trim Galore. After trimming, reads were aligned to Human Genome build hg38 using Bismark [21].

DNA methylation calling was then performed using the Bismark methylation calling tool. In short, the methylation value for a particular cytosine was calculated as number of reads that detect this cytosine in a methylated state divided by the total number of reads for this cytosine. CpG sites with a mean total

coverage of less than 10 reads per cytosine and CpG sites located on the Y chromosome were excluded from further analyses. Only CpG methylation on the forward strand was considered in this project. Methylation profiles were then smoothed and differentially methylated regions (DMRs) called using the BSmooth algorithm from Bioconductor bsseq package [22].

The DMR calling between islets from donors with T2D versus controls was performed with t-statistics cutoff of 1st and 99th percentile. DMRs were defined as regions of at least 3 consecutive CpG sites showing a significant methylation difference between the two groups, with an average absolute methylation difference of ≥5%. This threshold was selected based on an average sequencing depth of approximately 20× for the forward strand. The maximum allowed distance between consecutive CpG sites within a DMR was set to 300 bp [22].

Infinium 450K array

500 ng genomic DNA from pancreatic islets of human donors (**Table 1**) was bisulfite converted with the EZ DNA methylation kit (Zymo Research). DNA methylation was analysed with the Infinium HumanMethylation450K BeadChip [7]. All samples passed GenomeStudio® quality control steps based on built in control probes for staining, hybridization, extension and specificity and displayed high quality bisulfite conversion efficiency with a signal intensity above 4000 [23]. Probes with a mean detection *P*-value > 0.01, as well as SNP probes, non-CpG probes and CpG-SNP probes (MAF > 0.1 based on dbSNP) [24] were removed from further analysis. Probes that hybridize to more than one genomic location with a 49 or 50 base pair match [25] were also excluded. In total 475,885 probes remained for further analysis. Background correction and quantile normalization were performed using the lumi package from Bioconductor [26]. Beta Mixture Quantile dilation Method (BMIQ) was applied to correct for the two different probe types on the array [27]. Genomic locations of target CpG sites were converted from genome build hg19 to build hg38 using liftover tool [28].

RNA-seq data analysis

For RNA library preparation the TruSeq kit (Illumina) was used and the libraries were sequenced on a HiSeq2000 platform as previously described (Fadista et al. 2014). RNA sequencing data in FASTQ format was analyzed using the Expectation-Maximization pipeline (RSEM; http://deweylab.biostat.wisc.edu/rsem/), version 1.2.14, with hg38 as the reference genome, and gene annotation from Gencode version 22 as the gene model. For every gene, the transcript with highest mean Transcript Per Million (TPM) value was selected. For downstream analysis, each TPM value was added an offset of 0.01 and $\log_2$ transformed. Transcripts with a TPM value less than 0.1 were considered not expressed. Additionally, the remaining expressed transcripts were split into 3 equally sized groups, denoting lowly expressed, medium expressed and highly expressed transcripts.

Genomic annotation

Genomic elements, such as transcription start sites, transcription end sites, exons and introns for 198,442 transcripts corresponding to 60,483 genes were extracted from GENCODE release 22 (GRCh38). Each DMR was annotated based on its position in relation to all transcripts above and hence, one DMR can have multiple annotations.

Pyrosequencing

DNA methylation for biological replication (**Table S1**) of a selected genomic region in the most significant *PDX1* DMR (chr13:27921804:27925104) was analyzed by pyrosequencing and the PyroMark Q96ID (Qiagen, Hilden, Germany). PCR and sequencing primers were designed using PyroMark Assay Design 2.0 (Qiagen, **Table S2**), and all procedures were performed according to recommended protocols and as previously reported [29].

Luciferase assays

Luciferase assays were performed as previously described [6]. In short, 2000 bp fragments of the *TMED6* and *KIF3A* promoters (**Table S2**) were inserted immediately upstream of the transcription start site of the CpG-free luciferase reporter vector (pCpGL-basic). Two different DNA methyltransferases (New England Biolabs, Ipswich, MA, USA) were then used to methylate the constructs; SssI for complete methylation of both constructs and either HpaII (*TMED6*) or HhaI (*KIF3A*) for partial methylation. SssI methylates all cytosine residues within the double-stranded dinucleotide recognition sequence CG, while HpaII and HhaI methylates the internal cytosine residue in CCGG and GCGC sequences, respectively. INS-1 832/13 β-cells were then co-transfected with 25 ng methylated or mock-methylated pCpGL-vector including either of the two promoter inserts together with 4 ng of pRL renilla luciferase control reporter vector (Promega, Madison, WI, USA). Firefly luciferase luminescence, as a value of transcriptional activity, was measured for each construct with the Dual-Luciferase® Reporter Assay System (Promega) and an Infinite® M200 PRO multiplate reader (Tecan Group Ltd., Männedorf, Switzerland). Cells transfected with an empty pCpGL-vector were used as background control for firefly luciferase results, and untransfected cells were used as a background for renilla results.

Insulin secretion in β-cell lines

The INS-1 832/13 rat β-cell line [30] was used in functional knockdown and overexpression experiments as previously described [2]. For overexpression experiments, rat cDNA for *Nr4a3*, *Pid1*, and *Socs2* (sequences can be found in **Table S2**), with and without a c-terminal HA-tag, were inserted into the pcDNA3.1 expression plasmid. Overexpression at the RNA level was analysed with qPCR and assays against *Nr4a3* (Rn01354012_m1), *Pid1* (Rn01769975_m1), and *Socs2* (Rn00589521_m1). An assay for *Hprt1* (Rn01527840_m1) was used as an endogenous control. Overexpression at the protein level was determined by Western blot and primary antibodies against the HA-tag and actin as described [2]. The siRNA used for knockdown in INS-1 832/13 rat β-cells was s132780 (siPark2) and a negative control siRNA (siNC, Silencer Negative Control No.2, #AM4637, Thermo Fisher Scientific, Waltham, MA, USA). Knockdown was verified with a qPCR assay for *Park2* (Rn00571787_m1), and the above mentioned endogenous control. Insulin secretion was analyzed as previously described [31], except that

insulin was determined with ELISA (Mercodia, Uppsala, Sweden). Additionally, *SLC2A2* was silenced with siRNA s12928 (siSLC2A2) and a non-targeting siRNA was used as negative control (siNC, #AM4637, Ambion) in the human EndoC-βH1 cell line [32] using the Lipofectamine RNAiMax transfection reagent (Thermo Fisher Scientific) according to manufacturer's recommendation. Knockdown was verified with qPCR assays for *SLC2A2* (Hs01096905_m1) and *HPRT1* (4326321E). Insulin secretion experiments with the EndoC-βH1 cells were performed as previously described [32]. All siRNA and TaqMan assays were ordered from Thermo Fisher Scientific.

Statistical analysis

Donor characteristics, biological replication using pyrosequencing and average DNA methylation between T2D and control islets, as well as Luciferase experiments, were analyzed using unpaired t-tests. qPCR and insulin secretion experiments performed in β-cells were analyzed using paired t-tests. Average DNA methylation between transcripts of different expression levels was analyzed using ANOVA. All data are presented as mean ± SEM. The KEGG pathway analysis was performed using WebGestalt [33] with *P*-values adjusted for multiple testing according to the FDR method developed by Benjamini and Hochberg [34].

Statistical enrichment of the number of overlapping genomic regions between two features was calculated using null distribution obtained from 1,000,000 permutations of one of the features. Permutations were generated by randomly and uniformly selecting a new genomic position regardless of chromosome for the start of each region, while preserving the region length [35]. The permutation testing procedure was implemented in C++. All other statistical computations of the DNA methylation data were performed using R software (R Core Team, 2015).

Images

Heatmap plots were generated with R gplots package using Euclidian distance measure. Principal component analysis plots were produced using R ggbiplot package (https://github.com/vqv/ggbiplot). Other images were produced using ggplot2 package [36] and GraphPad prism.

Results

WGBS in human pancreatic islets

To characterize the methylome in human pancreatic islets, we generated WGBS data at a single-base resolution from pancreatic islets of 14 human donors, including eight normoglycemic donors and six donors diagnosed with T2D. The donor characteristics are described in **Table 1**. After preprocessing, an average of 74% of the resulting reads per sample were uniquely mapped to the human reference genome (hg38). Sequencing information and alignment statistics for each sample are reported in **Table S3**. The islet samples were sequenced with an average coverage of 21× per base and methylation levels of ~24x10^6 CpG sites (~83% of all CpG sites in the human genome) on the forward strand were obtained for all samples.

We compared the methylation data obtained by WGBS with data generated with the Infinium 450K array for the same 14 samples. The replicates of each islet sample analyzed by both WGBS and microarray showed high reproducibility ($r \geq 0.927$; $P < 2.2*10^{-16}$, **Table S3**).

Additionally, one islet sample was analyzed by WGBS using both Illumina type 3 and type 4 chemistry as well as by two different library preparation kits; the TruSeq (EpiGnome) DNA Methylation Kit and the NEXTflex™ Bisulfite Library Prep Kit. The high correlation between the WGBS data generated with the two different sequencing chemistries ($r = 0.995$) as well as with two different library preparations ($r = 0.987$) further confirmed the quality of our methylation data (**Fig S1A**).

In an unsupervised principal component analysis (PCA) of the islet WGBS data, the methylation data segregated according to sex (**Fig S1B)**, which is in agreement with our published Infinium 450K array data [31]. Next, we correlated the top five principal components of the WGBS data with T2D, age, sex and BMI. Here, T2D and sex correlated with one of the top five principal components ($P < 0.002$, **Table 2).**

Table 2. *P*-values for correlations of the top five principal components for the WGBS data in human pancreatic islets with type 2 diabetes (T2D), age, sex and BMI.

Principal component	T2D	Age (years)	Sex (m/f)	BMI (kg/m^2)
1	0.70	0.54	0.65	0.26
2	0.90	0.13	**0.0017**	0.47
3	0.25	0.19	0.41	0.62
4	0.07	0.58	0.75	0.11
5	**0.0019**	0.44	0.95	0.22

Differentially methylated regions in human pancreatic islets from donors with T2D

In order to address the epigenetic basis of T2D, we used BSmooth [22] to analyze the islet WGBS data from eight normoglycemic controls and six type 2 diabetic donors. This is an algorithm designed to determine DNA methylation in WGBS data and identify DMRs that account for biological variability. In our study, DMRs were defined as regions of three or more consecutive differentially methylated CpG sites with an average absolute methylation difference equal to or bigger than 5% between the groups. Based on this analysis, we identified 25,820 DMRs (**Table S4**). These were included in an unsupervised hierarchical clustering analysis presented as a heatmap in **Fig 1A**, which shows distinction in methylation between diabetics and controls. 13,696 DMRs showed average increased and 12,124 decreased levels of DNA methylation in islets from donors with T2D. The mean DMR size was 414 bp (range 6 - 3411 bp), and the mean CpG site count in the DMRs was 8.7 (range 3 - 164). The maximum absolute difference in methylation for a DMR was 27.5% and the mean absolute difference in methylation for all DMRs was 6.3% when comparing islets from diabetic versus control donors (**Table S4**). Among the DMRs found to have the largest absolute differences in methylation were regions annotated to *ARX* and *TFAM* (**Fig 1B-C**), two genes encoding proteins with important roles in islet function [37, 38].

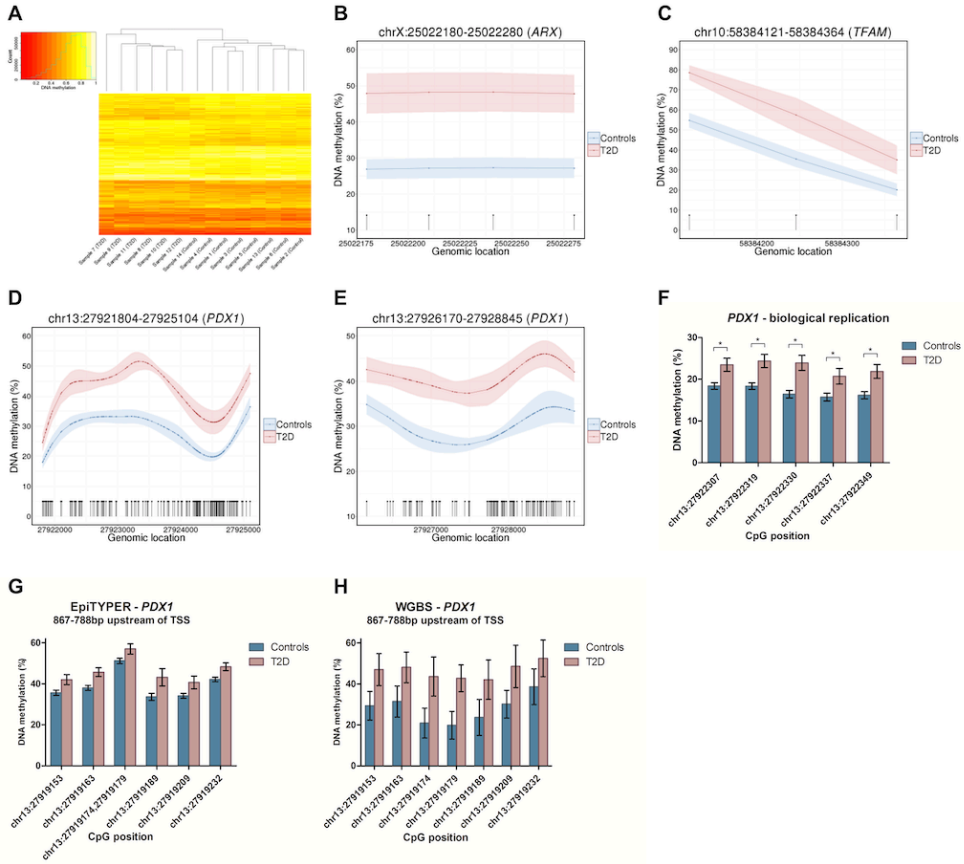


Fig 1. DMRs in human pancreatic islets from donors with T2D

A) Heatmap of the 25,820 T2D associated DMRs. Among the DMRs with the largest absolute difference were regions upstream of **B**) *ARX* and **C**) *TFAM*. Among the most significant DMRs were two large intergenic regions of *PDX1*, **D**) chr13:27921804:27925104 and **E**) chr13:27926170:27928845. **F**) Biological replication of the *PDX1* region presented in (**D**) in human pancreatic islets from an independent cohort of 56 normoglycemic control and 19 T2D donors. **G-H**) Validation of DNA methylation in the *PDX1* distal promoter in human pancreatic islets from T2D and control donors; **G**) Data from this study, based on WGBS DMR chr13:27918705-27919232 and **H**) Data from Yang et al. [6], produced using Sequenom's EpiTYPER technology. Data presented as mean ± SEM (* $P < 0.05$).

Interestingly, two of the most significant DMRs covered 164 and 105 CpG sites that span 3,301 and 2,676 bp intergenic regions of *PDX1* **(Fig 1D-E)**, a key transcription factor in pancreatic islets [39]. In total, seven DMRs were annotated to *PDX1* (**Table S4**). Using pyrosequencing, we could biologically replicate differential DNA methylation of CpG sites in the most significant DMR, located in an intronic region and exon two of *PDX1* (chr13:27921804-27925104, **Fig 1D, F**), in an independent cohort of 19 T2D and 56 control donors (**Table S1**). We previously reported increased DNA methylation of 10 CpG sites and decreased expression of *PDX1* in human pancreatic islets from donors with T2D [6]. Importantly, a DMR from the present study (chr13:27918705-27919232; ending 788 bp upstream of the *PDX1* TSS) confirmed our previous finding of differential methylation of *PDX1*, and comparisons of seven individual CpG sites covered by both studies validated the significant association with T2D (**Fig 1G**).

We next examined whether the T2D associated islet DMRs were located in the same region as the 65 known T2D candidate genes identified by the Diagram consortium [40]. Among our islet DMRs, 159 were annotated to 43 known T2D candidate genes (**Table S5)**. Out of these, the DMR with the largest mean absolute difference (11%) in methylation between T2D and control donors was located in the intron-exon boundary of *ADCY5*. Additionally, T2D candidate genes with the highest number of annotated DMRs were *GLIS3*, *THADA*, *KCNQ1* and *TCF7L2* (16, 14, 10 and 9 respectively).

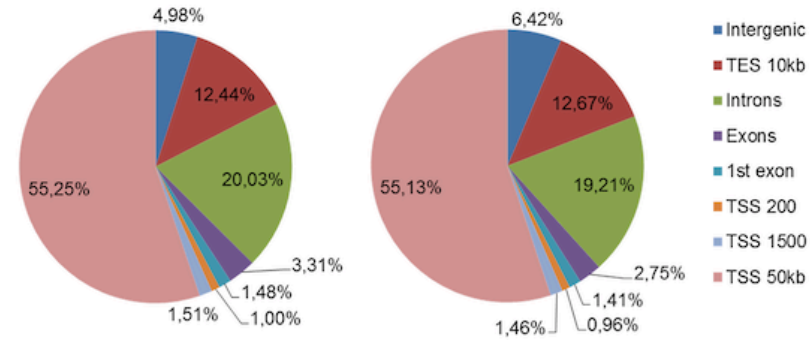
We further investigated if our islet DMRs covered any of the SNPs reported to be associated with T2D (www.genome.gov/gwastudies, accessed February 2nd, 2016). Among these SNPs, rs163184 was covered by a DMR annotated to *KCNQ1* (chr11:2825382:2826548) and rs11257655 by a DMR annotated to *RP11* (chr10:12265391:12266540).

The genomic distribution, chromatin state and transcription factor binding of T2D associated islet DMRs

We proceeded to examine the genomic distribution of the T2D associated DMRs in human pancreatic islets. We found a similar distribution of DMRs whether the average methylation level was increased or decreased in islets from donors with T2D, with ~55% of the DMRs located within TSS 50kb (1,501-

50,000bp upstream from the transcription start site), ~1.5% located within TSS 1,500 (201-1,500bp upstream from the transcription start site) and ~1.0% located within TSS 200 (1-200bp upstream from the transcription start site) (**Fig 2A**). For the intragenic regions, ~1.4% of all DMRs were located in the 1st exon, ~3.0% in the subsequent exons and ~20% in introns, whereas ~12.5% of the DMRs were located within TES 10kb (1-10,000bp downstream of the transcription end site). Additionally, ~5-6% of all DMRs were located >50 kb from the nearest transcript and considered intergenic.

A



B

Histone marks	# DMRs	Transcription factor binding sites	# DMRs
H3K4me3	3203 ↑	FOXA2	978 ↑
H3K9ac	3194 ↑	MAFB	311 ↑
H3K27ac	3744 ↑	NKX2.2	1943 ↑
H3K4me1	4935 ↑	NKX6.1	248 ↑
H3K27me3	1364 ↑	PDX1	624 ↑
H3K9me3	120 ↓	Enhancer regions	618 ↑
H3K36me3	581 ↓		

C

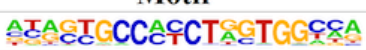
Motif	Name	P-value
	CTCF	1e-34
	FOXA2	1e-30
	NeuroD1	1e-21
	MAFA	1e-16
	RFX	1e-13
	PDX1	1e-6
	HNF1	1e-6

Fig 2. Genomic distribution of T2D associated DMRs

A) Genomic distribution of T2D associated islet DMRs, separated based on increased (*n*=13,696) or decreased (*n*=12,124) average DNA methylation. Each DMR can be annotated to several transcripts and

will then be counted in all gene regions, except from intergenic DMRs which only include DMRs located more than 50kb from a transcript and thereby not annotated to any of the TSS, intragenic or TES regions. **B**) Overlap between our T2D-associated DMRs and chromatin state, transcription factor binding and active enhancer regions. Arrows represent significant over- or underrepresentation ($P < 1*10^{-6}$). **C**) Enrichment of transcription factor recognition sequences in the T2D associated islet DMRs based on HOMER [41].

To explore the relationship between chromatin state and differential DNA methylation, we integrated all DMRs with genome-wide maps of histone modifications and enhancer regions previously generated in human pancreatic islets [18, 42]. When comparing our significant DMRs with epigenetic marks generated in human islets by the Roadmap Epigenomics Consortium [18], 12.4% of the T2D associated islet DMRs were occupied by modifications associated with active chromatin (3,203 by H3K4me3 and 3,194 by H3K9ac; **Table S6** and **Fig 2B**) which is a significant enrichment for both ($P < 1*10^{-6}$). Moreover, 14.5 and 19.1% of the T2D associated islet DMRs were occupied by histone modifications enriched at enhancer regions (3,744 by H3K27ac and 4,935 by H3K4me1, respectively), which is also significant enrichments ($P < 1*10^{-6}$). A smaller fraction of the DMRs were occupied by modifications associated with repressed chromatin (1,364 DMRs (5.3%) by H3K27me3 and 120 DMRs (0.5%) by H3K9me3). H3K27me3 was statistically overrepresented, while H3K9me3 was underrepresented ($P < 1*10^{-6}$) in this overlap. There was also an underrepresentation of DMRs (581 or 2.3%) overlapping with H3K36me3 ($P < 1*10^{-6}$). Additionally, 618 T2D associated islet DMRs overlapped with active enhancer regions of pancreatic islets identified by Pasquali et al. (**Table S6** and **Fig 2B**) [42], which is more than expected by chance ($P < 1*10^{-6}$).

We next used HOMER to test if any transcription factor recognition sequences were enriched in our significant islet DMRs [41]. Interestingly, motifs specific to key transcription factors in pancreatic islets, including FOXA2, NeuroD1, MAFA, RFX, PDX1 and HNF1 as well as binding sites for the insulator CCCTC binding factor (CTCF), were significantly enriched in the T2D associated islet DMRs (**Fig 2C, Table S7**). We also identified novel motifs with high statistical enrichment (**Table S7**).

To gain further insight into the relationship between the genomic binding of islet-specific transcription factors and differential DNA methylation in islets from subjects with T2D, we used ChIP-sequencing data for five key transcription factors (FOXA2, MAFB, NKX2.2, NKX6.1 and PDX1) in human pancreatic islets from Pasquali et al. [42]. We found an overrepresentation of DMRs overlapping with the binding sites of each of these transcription factors ($P < 1*10^{-6}$; **Table S8**, **Fig 2B**). Indeed, 978 (3.8%) of the T2D associated DMRs overlapped with genomic sites bound by FOXA2 and 311 (1.2%) DMRs overlapped with sites bound by MAFB. Additionally, 1943 (7.5%) DMRs overlapped with sites bound by NKX2.2, while 248 and 624 DMRs (1.0% and 2.4%) overlapped with sites bound by NKX6.1 and PDX1, respectively. Interestingly, we found an overlap between T2D associated DMRs annotated to *SLC2A2, KCNJ11* and *PDX1* and sites bound by PDX1 (**Table S8**).

Together, these data suggest that differential DNA methylation in transcription factor binding sites may be of importance in T2D.

Tissue specific DMRs

To identify DMRs from human pancreatic islets which are also altered between other tissues or cell types we analyzed the overlap between our T2D associated islet DMRs and a set of 716,087 cross-tissue dynamic DMRs identified by Ziller et al. [43]. Out of 25,820 T2D associated islet DMRs, 12,911 (49.8%) overlapped with a dynamic DMR identified in other tissues ($P < 1*10^{-6}$) (**Table S9**).

Altered expression of genes annotated to T2D associated DMRs

To determine whether the genes annotated to T2D associated DMRs also show altered expression in pancreatic islets of type 2 diabetic versus normoglycemic control donors, we combined the 25,820 DMRs presented in **Table S4** with RNA sequencing data from pancreatic islets of a previous publication [20]. Here, we identified 457 genes that had both significantly altered expression (False discovery rate < 5%; $q < 0.05$) and significant DMR(s) in pancreatic islets from donors with T2D (**Table S10**). Of note, these include genes of importance in islet function and metabolism such as *CACNA1D*, *CHL1*, *GLP1R*,

IGF1R, *IL6*, *NR4A3, PARK2*, *PDX1*, *PID1*, *SEPT9*, *SIK2*, *SLC2A2* (also known as *GLUT2*), *SOCS2* and *SOX6* [2, 39, 44-55] (**Fig 3A**).

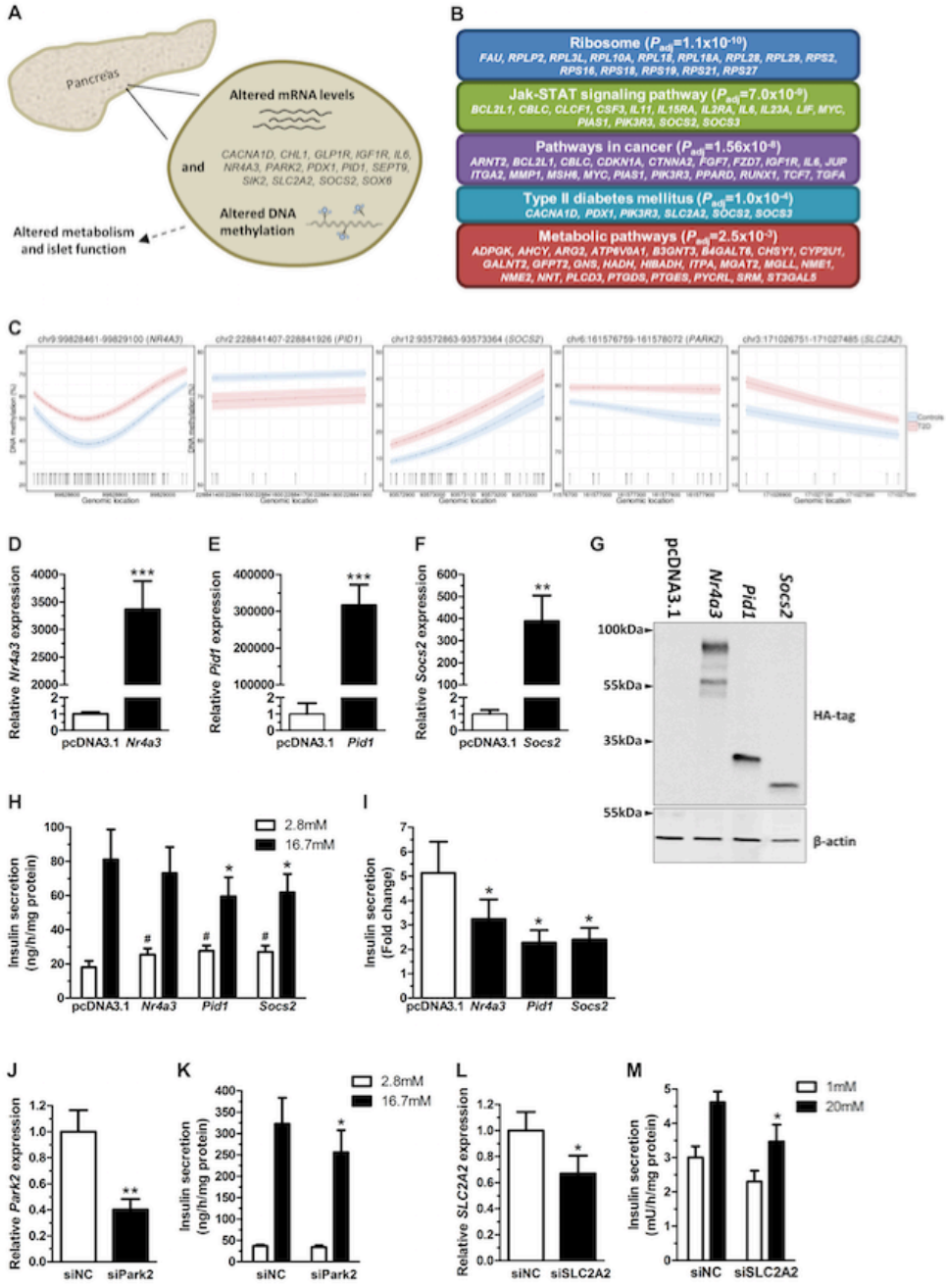


Fig 3. Altered expression, T2D associated DMRs and functional consequences in β-cells

In total, 457 genes had both significantly altered expression in pancreatic islets from donors with T2D ($q < 0.05$; [20]) and significant DMR(s) (**Table S10**). These include e.g. genes of importance in islet function and metabolism as depicted in **Fig A**. **B**) Selected KEGG pathways based on enrichment of genes annotated to T2D-associated DMRs that also show altered RNA expression in human pancreatic islets from donors with T2D [20]. 457 genes were included in the pathway analysis and the boxes include genes contributing to the enrichment score for each pathway. A full list of significantly enriched KEGG pathways are found in **Table S11. C)** DMRs from the five genes selected for functional studies; *NR4A3*, *PID1*, *SOCS2*, *PARK2* and *SLC2A2* (*GLUT2*)**.** Overexpression of *Nr4a3*, *Pid1*, and *Socs2* in INS-1 832/13 cells was verified by qPCR (**D-F**) and western blot (**G**). *** $P < 0.001$ and ** $P < 0.01$ as analysed by one-tailed paired t-tests (n=7). **H**) Overexpression of *Nr4a3*, *Pid1*, and *Socs2* resulted in perturbed insulin secretion (* $P < 0.05$ compared to siNC at 16.7mM glucose, and $^{\#}$ $P < 0.01$ compared to siNC at 2.8mM glucose, as analysed by two-tailed paired t-tests; n=7) and **I**) altered fold change of insulin secretion for each of the three genes. **J**) siRNA mediated knockdown of *Park2* was verified by qPCR (** $P < 0.01$ as analysed by a one-tailed paired t-test; n=6). **K**) *Park2* deficiency resulted in reduced insulin secretion at stimulatory glucose levels (* $P < 0.05$ as analysed by a two-tailed paired t-test; n=6). **L)** siRNA mediated knockdown of *GLUT2* in EndoC-βH1 was verified by qPCR (* $P < 0.01$ as analysed by a one-tailed paired t-test; n=4). **M)** *GLUT2* deficiency resulted in reduced insulin secretion at stimulatory glucose levels (* $P < 0.05$ as analysed by a two-tailed paired t-test; n=4).

Next, we performed a KEGG pathway analysis using WebGestalt [33] to identify cellular components and biological pathways with enrichment of genes that had both significantly altered expression and DMRs in islets from donors with T2D (n=457). Interestingly, these genes are significantly enriched for gene ontology (GO) categories including Ribosome, Jak-STAT signaling pathway, Pathways in cancer, Type II diabetes mellitus and Metabolic pathways (adjusted $P < 0.006$; **Fig 3B** and **Table S11**). These data further support that epigenetic and transcriptional changes in pancreatic islets may contribute to altered metabolism and T2D.

We continued to functionally study the impact of altered DNA methylation on the transcriptional activity using luciferase assays. For this experiment, we selected *TMED6* and *KIF3A*, two genes where the T2D

associated DMRs cover promoter regions and with an inverse relation between DNA methylation and mRNA expression (**Fig S1C, Table S10**). Altered *TMED6* expression has also been linked to reduced glucose-stimulated insulin secretion [20]. Luciferase constructs containing respective promoter sequence were either methylated or mock-methylated and transfected into clonal β-cells. In line with our human islet data, we found that either complete or partial methylation of the *KIF3A* promoter resulted in almost total shutdown of transcription. The same was seen for complete methylation of the *TMED6* promoter, while partial methylation did not affect transcriptional activity (**Fig S1D-E**).

We then asked if genes with both DMRs and altered expression in islets from subjects with T2D have a functional role in pancreatic β-cells. Genes were selected for functional follow-up experiments based on having multiple and/or large DMRs and exhibit differential expression ($q < 0.05$) in human diabetic islets (**Table S10; Fig 3A,C**) together with their potential role in β-cell function based on previous studies [44, 48, 50, 54]. To model the situation in humans with T2D, we overexpressed *Nr4a3*, *Pid1* and *Socs2* and silenced *Park2* in rat clonal 832/13 INS-1 β-cells, the most well-characterized β-cell line regarding insulin secretion [30] (**Fig 3D-G, J**). We then measured insulin secretion at basal (2.8mM) and stimulatory (16.7mM) glucose levels. While overexpression of *Pid1* and *Socs2* resulted in reduced glucose-stimulated insulin secretion, all three overexpressed genes caused a slight increase in basal insulin secretion (**Fig 3H**). These changes resulted in decreased fold change of insulin secretion (secretion at stimulatory divided by the secretion at basal glucose levels) in β-cells overexpressing either of the three genes (**Fig 3I**), which is in line with what is seen in subjects with T2D [2]. *Park2* deficiency resulted in reduced glucose-stimulated insulin secretion (**Fig 3J-K**). For similar reasons as above we also silenced *SLC2A2* (encoding GLUT2; **Fig 3A, C, L**) but here we used a human β-cell line (EndoC-βH1) since GLUT2 previously was found to be the key glucose transporter in rodent β-cells, whereas its role in human β-cells has been questioned [55-57]. Interestingly, reduced expression of *SLC2A2* resulted in impaired glucose-stimulated insulin secretion in human β-cells (**Fig 3L-M**). In no case were the identified secretory defects due to changes in insulin content (data not shown).

These findings support the notion that epigenetic modifications in human pancreatic islets may impact expression of genes that affect insulin secretion and potentially contribute to T2D.

The DNA methylome of human pancreatic islets

We finally sought to characterize the overall variability of the methylome in human pancreatic islets. We found the degree of DNA methylation throughout the genome to be highly correlated among the analyzed islets samples (r = 0.973 – 0.989) and the average level of DNA methylation was 75.9%. The distribution of the DNA methylation level in human islets is bimodal with the highest peak at 90.2%, showing that most CpG sites are highly methylated, whereas the second highest peak is seen at 1.4%, representing CpG sites with no or low levels of DNA methylation (**Fig 4A**).

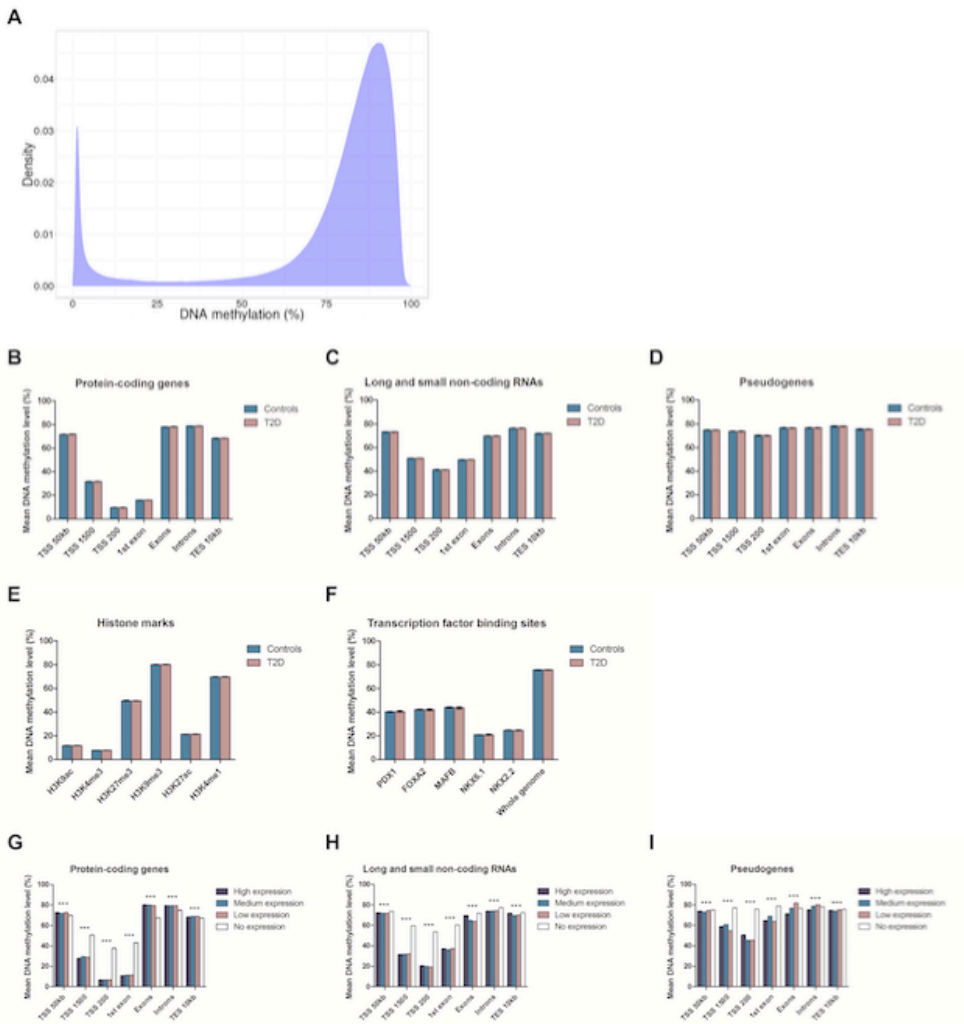


Fig 4. The DNA methylome of human pancreatic islets

A) The distribution of DNA methylation in human islets. **B-D**) Average DNA methylation in human pancreatic islets, separated by different transcript types (protein-coding, non-coding and pseudogenes) and gene regions. Here, TSS 50kb represents 1,501-50,000bp upstream from the transcription start site, TSS 1500 represents 201-1,500bp upstream from the transcription start site, TSS 200 represents 1-200bp upstream from the transcription start site and TES 10kb represents 1-10,000bp downstream of the transcription end site. T2D and normoglycemic controls display no differences in genome-wide average methylation for any genomic region or transcript type (P = 0.4-1.0). **E-F**) Average DNA methylation for regions overlapping with different histone marks or transcription factor binding sites. **G-I**) Average methylation levels are significantly different between transcripts of different expression levels. Additionally, different genomic regions display specific methylation patterns which are also dependent on transcript type. Data presented as mean ± SEM (***ANOVA $P < 0.0001$).

Next, we computed the average genome-wide methylation level in relation to different genomic regions in the human islets. While introns and exons had the overall highest degree of methylation (78.5% and 77.4%), regions close to the TSS such as the 1st exon and promoter regions (TSS 200 and TSS 1500, defined as regions 200bp and 1500bp upstream of TSS) had the lowest degree of methylation (34.7%, 25.4% and 44.4%, respectively). Additionally, regions more distant to the TSS had a rather high methylation level (TSS 50kb 73.0% and TES 10kb 71.2%).

We also studied methylation based on annotation to different types of transcripts; protein-coding genes, non-coding RNAs and pseudogenes (**Fig 4B-D**). Interestingly, there was a striking difference in average DNA methylation between the different transcript types, where the typical drop in DNA methylation seen close to the TSS for protein-coding genes (TSS 1500, TSS 200 and 1st exon) was less pronounced in non-coding RNAs (long and small combined) and almost completely absent in pseudogenes (**Fig 4B-D**). Additionally, there were no differences in average DNA methylation for either genomic regions or transcript types between islet DNA from T2D compared to normoglycemic control donors (**Fig 4B-D**; P = 0.4-1.0).

To gain further insight into how the islet epigenome is coordinated, we studied the relationship between DNA methylation levels and histone modifications across the genome in human pancreatic islets. Here, we integrated our islet WGBS data with epigenetic marks generated in human islets by the Roadmap Epigenomics Consortium [18]. Regions occupied by histone modifications associated with active chromatin had the lowest degree of methylation (11.8% for H3K9ac and 7.8% for H3K4me3), while regions occupied by modifications associated with repressive chromatin had a higher methylation level (49.6% for H3K27me3 and 80.2% for H3K9me3) (**Fig 4E**). Additionally, regions occupied by histone modifications considered to be enriched at enhancer regions had either quite a low (21.4% for H3K27ac) or a high (69.8% for H3K4me1) degree of methylation (**Fig 4E**). We also examined the relationship between DNA methylation levels and genomic binding of islet-specific transcription factors (PDX1, FOXA2, MAFB, NKX6.1 and NKX2.2), by combining our genome-wide methylation data with ChIP-seq data generated by Pasquali et al. [42]. Regions occupied by islet-specific transcription factors had a lower degree of methylation (range 20.9% - 43.9%) compared to the whole genome (**Fig 4F**). There were no differences in the degree of DNA methylation in any of the regions occupied by the studied histone marks or transcription factors between islet DNA from T2D compared to normoglycemic control donors (**Fig 4E-F**; $P = 0.4\text{-}1.0$).

We proceeded to explore the relationship between DNA methylation and gene expression levels using WGBS and RNA sequencing data from the same 14 human islets. Here, we categorized 60,483 transcripts into not expressed (Transcript Per Million (TPM) < 0.1, 38,261 transcripts in total), and equally sized groups of lowly (7,407), medium (7,407) and highly expressed (7,408) transcripts. We also studied protein-coding genes, non-coding RNAs and pseudogenes separately. We found an association between DNA methylation and the expression level in all genomic regions and for all types of transcripts ($P < 0.0001$) (**Fig 4G-I**). Importantly, protein-coding genes not expressed were hypermethylated in regions close to the TSS (TSS 1500, TSS 200 and 1st exon; **Fig 4G**). On the other hand, protein-coding non-transcribed genes had significantly lower methylation levels in the exon and intron regions compared with the transcribed genes (**Fig 4G**), supporting that increased methylation in the gene body is associated with a higher level of gene transcription [58]. However, more modest differences in DNA methylation were seen in the genomic regions when comparing the different

expression levels of transcribed genes, i.e. dividing transcribed genes in lowly, medium and highly expressed genes (**Fig 4G-I**). Some similarities in the methylation pattern were seen for both non-coding RNAs and pseudogenes compared with protein-coding genes (**Fig 4H-I**). However, regions close to the TSS were hypermethylated in non-transcribed genes and the overall degree of methylation in these regions was much higher in the transcribed non-coding RNAs and pseudogenes (**Fig 4H-I**).

Discussion

Alternative approaches are needed to put an end to the rapid increase in T2D incidence. This study provides valuable insights into T2D pathology and human biology, including the first comprehensive and unbiased DNA methylation analysis of 24 million CpG sites in a case control-cohort of human pancreatic islets. By this approach, we identified 25,820 DMRs in islets from diabetic and control donors. These cover loci annotated to genes with known functions in islets and T2D pathogenesis as well as novel candidate genes. Interestingly, the identified DMRs were significantly enriched in both experimentally identified and putative binding motifs for islet-specific transcription factors. Integrating our WGBS data with RNA-seq data further identified novel candidate genes that contribute to islet dysfunction and impaired insulin secretion.

Intriguingly, seven of the identified DMRs were annotated to *PDX1*, which encodes a transcription factor of key importance during pancreatic development as well as in mature β-cells where it e.g. regulates expression of the insulin gene and several other genes of importance for β-cell function [59]. Additionally, mutations in *PDX1* cause MODY4 in humans [60] and knockout of *Pdx1* in rodent β-cells causes diabetes [61]. Using a candidate gene approach, we previously found increased DNA methylation and decreased expression of *PDX1* in islets from subjects with T2D [6]. One DMR annotated to *PDX1* in the present study validated these previous data. Epigenetic changes of *Pdx1* in islets of rats exposed to an impaired intrauterine environment predispose to diabetes and islet dysfunction in adult life [62, 63]. We have also previously shown that high glucose levels directly increase DNA methylation of *Pdx1* in clonal β-cells cultured *in vitro* [6] and that a SNP in *PDX1* associated with hyperglycemia also alters methylation of *PDX1* in human islets [64]. Together, these data support that epigenetic modifications of *PDX1* may contribute to the pathogenesis of diabetes.

The majority of SNPs associated with T2D impact insulin secretion rather than insulin action, pointing to importance of T2D candidate genes in pancreatic islet function [1, 65, 66]. However, these SNPs only explain a modest proportion of the estimated heritability of T2D and it is possible that combinations of genetic and epigenetic variation contribute to disease susceptibility [64, 67]. This theory is supported by the fact that a large number of our T2D associated DMRs are located in the same regions as T2D

candidate genes identified by GWAS. For example, we found islet DMRs annotated to *TCF7L2*, *ADCY5*, *KCNQ1* and *GLIS3*. Also functional studies show the importance of several of these candidate genes, such as *TCF7L2* and *ADCY5*, in β-cells and pancreatic islets [68, 69], further underpinning the role of epigenetic dysregulation in the etiology of T2D.

Previous studies of human pancreatic islets from donors with and without T2D have only investigated DNA methylation of selected candidate genes or a small fraction of genomic CpG sites [2-6, 46]. The current study engaged an approach to locate continuous regions of differential CpG methylation, i.e. DMRs [22]. When comparing our WGBS results with the 450K array data within samples, we observed correlations from 0.93 to 0.94. This indicates that our WGBS data is of high quality, in line with two previous studies in other tissues [70, 71]. It has been suggested that minimal sequencing requirements in WGBS experiments starts from 5× [72]. However, it was also stated that higher coverage is required to detect shorter DMRs with smaller methylation differences. Based on our previous findings, we expected some CpG sites to show absolute differences in methylation less than 10% between islets from type 2 diabetic and control donors [2], hence we implemented a higher coverage (21×) in our study design. Indeed, we discovered numerous DMRs that show absolute methylation differences between 5 and 10%. This would require at least 10× coverage for detection, which was also the level we used for cutoff in our bioinformatic analysisis. As the epigenetic process is highly dynamic both over time and in different tissues and cell types, future studies should aim to distinguish between potentially informative and non-informative genomic regions to reduce cost as well as amount of data to process. For example, DMRs shown to be highly variable between tissues may be of significance in determining the characteristic of specific cell types [43], but may also be regions more prone to epigenetic alterations through e.g. environmental influences or disease. Intriguingly, our results support this hypothesis since approximatelt 50% of our T2D associated islet DMRs overlapped with dynamic DMRs identified in other tissues [43]. In addition, DMRs overlapping with enhancer regions and transcription factor binding sites may also have an important role in gene regulation and disease development [73, 74]. Indeed, these regions were overrepresented in the DMRs detected in our study. A recent study by Domcke et al proposed competition between DNA methylation and binding of transcription factors to DNA [74]. In relation to their data, it is worth mentioning that we found 20.9-43.9% methylation in regions occupied

by islet specific transcription factors such as PDX1, FOXA2, MAFB, NKX6.1 and NKX2.2. It is possible that these transcription factors bind to DNA in a subset of islet cells with hypomethylated DNA, while they may not bind to DNA in the cells with methylated DNA. Methylation may thereby regulate gene expression differentially in different islet cells. The fact that motifs for several islet specific transcription factors were enriched in T2D associated islet DMRs further support an important role of methylation in these regulatory regions.

When combining the identified DMRs with RNA-seq data from islets of T2D and control donors, we identified 457 genes with both altered methylation and expression. One example is *CDKN1A*, which we previously have shown is epigenetically altered in T2D islets and influences insulin secretion in clonal β-cells [2]. Interestingly, we also found increased DNA methylation and decreased expression of *SLC2A2* in islets from subjects with T2D. *SLC2A2* encodes GLUT2, which is the major glucose transporter in rodent islets, whereas other glucose transporters have been suggested to be more important in human β-cells [56]. Nevertheless, Sansbury et al. reported a homozygous loss of function mutation in *SLC2A2* as a rare cause of neonatal diabetes, suggesting a potential role for GLUT2 in human β-cells [57]. Hence, the data by Sansbury et al. together with our epigenetic data encouraged us to study the role of GLUT2 in human β-cells. Importantly, silencing the expression of *SLC2A2* resulted in decreased glucose-stimulated insulin secretion in a human β-cell line supporting a role for GLUT2 as a glucose transporter not only in rodent but also in human β-cells.

To further model the situation in human T2D, we performed functional follow-up experiments of additional genes exhibiting significant DMRs and altered expression in islets from diabetic donors. Here, we overexpressed *Nr4a3*, which encodes a nuclear receptor that reduces insulin gene expression by modulating the expression of *Pdx1* and *NeuroD1* [75], *Socs2*, which encodes a suppressor of cytokine signalling found to regulate proinsulin processing and insulin secretion in transgenic mice [50] and *Pid1*, which encodes phosphotyrosine interaction domain-containing protein 1 that has been implicated in mitochondrial dysfunction and insulin resistance [44]. Interestingly, overexpression of all these three genes decreased the ratio between glucose-stimulated and basal insulin secretion, which is in line with what is seen in human diabetic islets. We also silenced the expression of *Park2*, which encodes parkin and has been shown to regulate the mitochondrial control system in β-cells [48]. Again, this impaired

glucose-stimulated insulin secretion. These experiments in clonal β-cells support that genes identified by WGBS and RNA-seq may contribute to islet dysfunction in subjects with T2D.

DNA methylation was initially thought to be a silencing mark, however more recent data show that its function may vary with genomic context and the situation is more complex than initially thought [58]. Here, we integrated the islet methylome with histone modifications analysed by the Roadmap Epigenomics Consortium in human pancreatic islets [18]. Notably, Kundaje et al. generated comprehensive WGBS data in several tissues, but only reduced representation bisulfite sequencing methylation data for human pancreatic islets [18] and our study is hence the first to provide WGBS coverage in human islets. In line with what may be expected, histone marks enriched around the TSS of actively transcribed genes (e.g. H3K9ac, H3K27ac and H3K4me3) as well as active enhancers (e.g. H3K27ac) had a relatively low degree of DNA methylation, while histone marks enriched around inactive TSS (e.g. H3K27me3) or inactive regions (e.g. H3K9me3) had a medium or high degree of DNA methylation. We also found a relatively high degree of DNA methylation in regions enriched with H3K4me1, which is a modification found both at enhancer regions and gene bodies of actively transcribed genes. Thus the high degree of DNA methylation in regions enriched for H3K4me1 is in line with the high degree of DNA methylation that is often found in gene bodies [76-79].

We also integrated the full islet methylome and transcriptome and examined the degree of methylation in different genomic regions of protein-coding genes, non-coding RNAs and pseudogenes. It is remarkable that DNA methylation of the different transcript types shows a completely different pattern, highlighting the need to take this into account when determining activity of a DNA region based on the epigenetic profile. The relation to gene expression was also distinct when comparing non-expressed and expressed transcripts, with reduced impact with increasing distance from TSS. The complex relation between WGBS methylation and RNA-seq data has been shown also in a few other tissues, supporting methylation to have a gene regulator role also outside promoters [15, 16]. Taken together, our data clearly show that the relationship between DNA methylation and gene expression depends on transcript type, expression level and distance from the TSS.

As DMRs may be cell-type or tissue specific [43], it is important to minimize the impact of cellular heterogeneity within the study. Here all samples were selected to have a high purity, i.e. a high endocrine

content, without differences between the groups. Furthermore, we have previously shown that there is no difference in β-cell content in the pancreatic islets from donors with T2D compared with controls [2]. These data support that the T2D associated DMRs we identify are not due to altered cell composition between the groups.

This comprehensive study identified novel diabetes-related changes in DNA methylation throughout the genome that support a central role for epigenetics in T2D. We need to combine multiple layers of biological information to get closer to understanding the pathogenesis of T2D and the progression of the disease. Here, we combined WGBS and RNA-seq data from human islets with known regulatory elements such as histone marks, transcription factor binding sites and enhancer regions. These integrated data advances our understanding of the etiology of T2D. Together, our results highlight the importance of epigenetic dysregulation in pancreatic islets and T2D pathogenesis.

Acknowledgements

Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Human pancreatic islets were obtained through collaboration with Olle Korsgren at the Nordic Network for Islet Transplantation (Uppsala University, Sweden) and the Human Tissue Lab at Lund Univeristy Diabetes Centre, coordinated by Ulrika Krus. We thank Anna-Maria Veljanovska-Ramsay, Lund University for technical assistance and Prof Manolis Kellis, MIT, Boston for support with the computational analysis.

This work was supported by grants from the Swedish Research Council, Region Skåne (ALF), Knut and Alice Wallenberg Foundation, Novo Nordisk Foundation, EFSD/Lilly, Söderberg Foundation, The Royal Physiographic Society in Lund, The Swedish Diabetes foundation, Påhlsson Foundation, EXODIAB and Linné grant (B31 5631/2006). Part of the work described in this manuscript was undertaken as part of the 2013-2014 BLUE ScY educational exchange program, which was supported by the Faculty of Medicine at Umeå University and EPiHealth (Lund University).

REFERENCES

1. Groop L, Pociot F. Genetics of diabetes--are we missing the genes or the disease? Molecular and cellular endocrinology. 2014;382(1):726-39. doi: 10.1016/j.mce.2013.04.002. PubMed PMID: 23587769.
2. Dayeh T, Volkov P, Salo S, Hall E, Nilsson E, Olsson AH, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. PLoS genetics. 2014;10(3):e1004160. doi: 10.1371/journal.pgen.1004160. PubMed PMID: 24603685; PubMed Central PMCID: PMC3945174.
3. Ling C, Del Guerra S, Lupi R, Rönn T, Granhall C, Luthman H, et al. Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia. 2008;51(4):615-22. Epub 2008/02/14. doi: 10.1007/s00125-007-0916-5. PubMed PMID: 18270681; PubMed Central PMCID: PMC2270364.
4. Volkmar M, Dedeurwaerder S, Cunha DA, Ndlovu MN, Defrance M, Deplus R, et al. DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. The EMBO journal. 2012;31(6):1405-26. Epub 2012/02/02. doi: 10.1038/emboj.2011.503. PubMed PMID: 22293752; PubMed Central PMCID: PMC3321176.
5. Yang BT, Dayeh TA, Kirkpatrick CL, Taneera J, Kumar R, Groop L, et al. Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. Diabetologia. 2011;54(2):360-7. Epub 2010/11/26. doi: 10.1007/s00125-010-1967-6. PubMed PMID: 21104225; PubMed Central PMCID: PMC3017313.
6. Yang BT, Dayeh TA, Volkov PA, Kirkpatrick CL, Malmgren S, Jing X, et al. Increased DNA Methylation and Decreased Expression of PDX-1 in Pancreatic Islets from Patients with Type 2 Diabetes. Mol Endocrinol. 2012. Epub 2012/05/10. doi: 10.1210/me.2012-1004. PubMed PMID: 22570331.
7. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288-95. Epub 2011/08/16. doi: S0888-7543(11)00180-7 [pii]
10.1016/j.ygeno.2011.07.007. PubMed PMID: 21839163.
8. Agirre X, Castellano G, Pascual M, Heath S, Kulis M, Segura V, et al. Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. Genome research. 2015;25(4):478-87. Epub 2015/02/04. doi: 10.1101/gr.180240.114. PubMed PMID: 25644835; PubMed Central PMCID: PMCPMC4381520.
9. Busche S, Shao X, Caron M, Kwan T, Allum F, Cheung WA, et al. Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation. Genome biology. 2015;16(1):290. doi: 10.1186/s13059-015-0856-1. PubMed PMID: 26699896; PubMed Central PMCID: PMCPMC4699357.
10. Court F, Tayama C, Romanelli V, Martin-Trujillo A, Iglesias-Platas I, Okamura K, et al. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome research. 2014;24(4):554-69. Epub 2014/01/10. doi: 10.1101/gr.164913.113. PubMed PMID: 24402520; PubMed Central PMCID: PMCPMC3975056.
11. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci U S A. 2012;109(26):10522-7. Epub 2012/06/13. doi: 10.1073/pnas.1120658109. PubMed PMID: 22689993; PubMed Central PMCID: PMCPMC3387108.
12. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, et al. Dynamic changes in the human methylome during differentiation. Genome research. 2010;20(3):320-31. Epub 2010/02/06. doi: 10.1101/gr.101907.109. PubMed PMID: 20133333; PubMed Central PMCID: PMCPMC2840979.
13. Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, et al. The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol. 2010;8(11):e1000533. Epub 2010/11/19. doi: 10.1371/journal.pbio.1000533. PubMed PMID: 21085693; PubMed Central PMCID: PMCPMC2976721.
14. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315-

22. Epub 2009/10/16. doi: 10.1038/nature08514. PubMed PMID: 19829295; PubMed Central PMCID: PMCPMC2857523.
15. Lou S, Lee HM, Qin H, Li JW, Gao Z, Liu X, et al. Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. Genome biology. 2014;15(7):408. Epub 2014/07/31. doi: 10.1186/s13059-014-0408-0. PubMed PMID: 25074712; PubMed Central PMCID: PMCPMC4189148.
16. Wen L, Li X, Yan L, Tan Y, Li R, Zhao Y, et al. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. Genome biology. 2014;15(3):R49. Epub 2014/03/07. doi: 10.1186/gb-2014-15-3-r49. PubMed PMID: 24594098; PubMed Central PMCID: PMCPMC4053808.
17. Zeng J, Konopka G, Hunt BG, Preuss TM, Geschwind D, Yi SV. Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. American journal of human genetics. 2012;91(3):455-65. Epub 2012/08/28. doi: 10.1016/j.ajhg.2012.07.024. PubMed PMID: 22922032; PubMed Central PMCID: PMCPMC3511995.
18. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317-30. Epub 2015/02/20. doi: 10.1038/nature14248. PubMed PMID: 25693563.
19. Olsson AH, Yang BT, Hall E, Taneera J, Salehi A, Nitert MD, et al. Decreased expression of genes involved in oxidative phosphorylation in human pancreatic islets from patients with type 2 diabetes. Eur J Endocrinol. 2011;165(4):589-95. doi: 10.1530/EJE-11-0282. PubMed PMID: 21775499; PubMed Central PMCID: PMCPMC3178933.
20. Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra JL, Taneera J, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A. 2014;111(38):13924-9. doi: 10.1073/pnas.1402665111. PubMed PMID: 25201977; PubMed Central PMCID: PMC4183326.
21. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27(11):1571-2. doi: 10.1093/bioinformatics/btr167. PubMed PMID: 21493656; PubMed Central PMCID: PMC3102221.
22. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome biology. 2012;13(10):R83. doi: 10.1186/gb-2012-13-10-r83. PubMed PMID: 23034175; PubMed Central PMCID: PMC3491411.
23. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One. 2009;4(12):e8274. Epub 2009/12/19. doi: 10.1371/journal.pone.0008274. PubMed PMID: 20019873; PubMed Central PMCID: PMC2793425.
24. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics & chromatin. 2013;6(1):4. doi: 10.1186/1756-8935-6-4. PubMed PMID: 23452981; PubMed Central PMCID: PMC3740789.
25. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics : official journal of the DNA Methylation Society. 2013;8(2). Epub 2013/01/15. PubMed PMID: 23314698.
26. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008;24(13):1547-8. Epub 2008/05/10. doi: 10.1093/bioinformatics/btn224. PubMed PMID: 18467348.
27. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189-96. Epub 2012/11/24. doi: 10.1093/bioinformatics/bts680. PubMed PMID: 23175756; PubMed Central PMCID: PMC3546795.
28. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, et al. The UCSC Genome Browser database: 2016 update. Nucleic acids research. 2016;44(D1):D717-25. doi: 10.1093/nar/gkv1275. PubMed PMID: 26590259; PubMed Central PMCID: PMCPMC4702902.
29. Rönn T, Volkov P, Davegardh C, Dayeh T, Hall E, Olsson AH, et al. A Six Months Exercise Intervention Influences the Genome-wide DNA Methylation Pattern in Human Adipose Tissue. PLoS

genetics. 2013;9(6):e1003572. Epub 2013/07/05. doi: 10.1371/journal.pgen.1003572. PubMed PMID: 23825961.
30. Hohmeier HE, Mulder H, Chen G, Henkel-Rieger R, Prentki M, Newgard CB. Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion. Diabetes. 2000;49(3):424-30. PubMed PMID: 10868964.
31. Hall E, Volkov P, Dayeh T, Esguerra JL, Salo S, Eliasson L, et al. Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. Genome biology. 2014;15(12):522. Epub 2014/12/18. doi: 10.1186/s13059-014-0522-z. PubMed PMID: 25517766; PubMed Central PMCID: PMCPMC4256841.
32. Ravassard P, Hazhouz Y, Pechberty S, Bricout-Neveu E, Armanet M, Czernichow P, et al. A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. J Clin Invest. 2011;121(9):3589-97. doi: 10.1172/JCI58447. PubMed PMID: 21865645; PubMed Central PMCID: PMCPMC3163974.
33. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic acids research. 2005;33(Web Server issue):W741-8. doi: 10.1093/nar/gki475. PubMed PMID: 15980575; PubMed Central PMCID: PMCPMC1160236.
34. Benjamini, Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B. 1995;57:289-300.
35. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform. 2015;16(1):59-70. doi: 10.1093/bib/bbt086. PubMed PMID: 24300110; PubMed Central PMCID: PMCPMC4293378.
36. Hadley, Wickham. ggplot2. 1 ed: Springer-Verlag New York; 2009.
37. Collombat P, Hecksher-Sorensen J, Krull J, Berger J, Riedel D, Herrera PL, et al. Embryonic endocrine pancreas and mature beta cells acquire alpha and PP cell phenotypes upon Arx misexpression. J Clin Invest. 2007;117(4):961-70. doi: 10.1172/JCI29115. PubMed PMID: 17404619; PubMed Central PMCID: PMCPMC1839241.
38. Gauthier BR, Wiederkehr A, Baquie M, Dai C, Powers AC, Kerr-Conte J, et al. PDX1 deficiency causes mitochondrial dysfunction and defective insulin secretion through TFAM suppression. Cell Metab. 2009;10(2):110-8. doi: 10.1016/j.cmet.2009.07.002. PubMed PMID: 19656489; PubMed Central PMCID: PMCPMC4012862.
39. Kaneto H, Miyatsuka T, Kawamori D, Yamamoto K, Kato K, Shiraiwa T, et al. PDX-1 and MafA play a crucial role in pancreatic beta-cell differentiation and maintenance of mature beta-cell function. Endocr J. 2008;55(2):235-52. PubMed PMID: 17938503.
40. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44(9):981-90. Epub 2012/08/14. doi: 10.1038/ng.2383. PubMed PMID: 22885922; PubMed Central PMCID: PMC3442244.
41. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004. PubMed PMID: 20513432; PubMed Central PMCID: PMCPMC2898526.
42. Pasquali L, Gaulton KJ, Rodriguez-Segui SA, Mularoni L, Miguel-Escalada I, Akerman I, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet. 2014;46(2):136-43. doi: 10.1038/ng.2870. PubMed PMID: 24413736; PubMed Central PMCID: PMCPMC3935450.
43. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500(7463):477-81. doi: 10.1038/nature12433. PubMed PMID: 23925113; PubMed Central PMCID: PMC3821869.
44. Bonala S, McFarlane C, Ang J, Lim R, Lee M, Chua H, et al. Pid1 induces insulin resistance in both human and mouse skeletal muscle during obesity. Mol Endocrinol. 2013;27(9):1518-35. doi: 10.1210/me.2013-1048. PubMed PMID: 23927930.
45. Ellingsgaard H, Hauselmann I, Schuler B, Habib AM, Baggio LL, Meier DT, et al. Interleukin-6 enhances insulin secretion by increasing glucagon-like peptide-1 secretion from L cells and alpha cells. Nat Med. 2011;17(11):1481-9. doi: 10.1038/nm.2513. PubMed PMID: 22037645; PubMed Central PMCID: PMCPMC4286294.

46. Hall E, Dayeh T, Kirkpatrick CL, Wollheim CB, Dekker Nitert M, Ling C. DNA methylation of the glucagon-like peptide 1 receptor (GLP1R) in human pancreatic islets. BMC medical genetics. 2013;14:76. Epub 2013/07/25. doi: 10.1186/1471-2350-14-76. PubMed PMID: 23879380; PubMed Central PMCID: PMC3727960.
47. Iguchi H, Urashima Y, Inagaki Y, Ikeda Y, Okamura M, Tanaka T, et al. SOX6 suppresses cyclin D1 promoter activity by interacting with beta-catenin and histone deacetylase 1, and its down-regulation induces pancreatic beta-cell proliferation. The Journal of biological chemistry. 2007;282(26):19052-61. doi: 10.1074/jbc.M700460200. PubMed PMID: 17412698.
48. Jin HS, Kim J, Lee SJ, Kim K, Go MJ, Lee JY, et al. The PARK2 gene is involved in the maintenance of pancreatic beta-cell functions related to insulin production and secretion. Molecular and cellular endocrinology. 2014;382(1):178-89. doi: 10.1016/j.mce.2013.09.031. PubMed PMID: 24096089.
49. Kulkarni RN, Holzenberger M, Shih DQ, Ozcan U, Stoffel M, Magnuson MA, et al. beta-cell-specific deletion of the Igf1 receptor leads to hyperinsulinemia and glucose intolerance but does not alter beta-cell mass. Nat Genet. 2002;31(1):111-5. doi: 10.1038/ng872. PubMed PMID: 11923875.
50. Lebrun P, Cognard E, Gontard P, Bellon-Paul R, Filloux C, Berthault MF, et al. The suppressor of cytokine signalling 2 (SOCS2) is a key repressor of insulin secretion. Diabetologia. 2010;53(9):1935-46. doi: 10.1007/s00125-010-1786-9. PubMed PMID: 20499047.
51. Reinbothe TM, Alkayyali S, Ahlqvist E, Tuomi T, Isomaa B, Lyssenko V, et al. The human L-type calcium channel Cav1.3 regulates insulin release and polymorphisms in CACNA1D associate with type 2 diabetes. Diabetologia. 2013;56(2):340-9. doi: 10.1007/s00125-012-2758-z. PubMed PMID: 23229155.
52. Sakamaki J, Fu A, Reeks C, Baird S, Depatie C, Al Azzabi M, et al. Role of the SIK2-p35-PJA2 complex in pancreatic beta-cell functional compensation. Nat Cell Biol. 2014;16(3):234-44. doi: 10.1038/ncb2919. PubMed PMID: 24561619; PubMed Central PMCID: PMCPMC4107453.
53. Taneera J, Lang S, Sharma A, Fadista J, Zhou Y, Ahlqvist E, et al. A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. Cell Metab. 2012;16(1):122-34. doi: 10.1016/j.cmet.2012.06.006. PubMed PMID: 22768844.
54. Tessem JS, Moss LG, Chao LC, Arlotto M, Lu D, Jensen MV, et al. Nkx6.1 regulates islet beta-cell proliferation via Nr4a1 and Nr4a3 nuclear receptors. Proc Natl Acad Sci U S A. 2014;111(14):5242-7. doi: 10.1073/pnas.1320953111. PubMed PMID: 24706823; PubMed Central PMCID: PMCPMC3986138.
55. van de Bunt M, Gloyn AL. A tale of two glucose transporters: how GLUT2 re-emerged as a contender for glucose transport into the human beta cell. Diabetologia. 2012;55(9):2312-5. doi: 10.1007/s00125-012-2612-3. PubMed PMID: 22696037.
56. De Vos A, Heimberg H, Quartier E, Huypens P, Bouwens L, Pipeleers D, et al. Human and rat beta cells differ in glucose transporter but not in glucokinase gene expression. J Clin Invest. 1995;96(5):2489-95. doi: 10.1172/JCI118308. PubMed PMID: 7593639; PubMed Central PMCID: PMCPMC185903.
57. Sansbury FH, Flanagan SE, Houghton JA, Shuixian Shen FL, Al-Senani AM, Habeb AM, et al. SLC2A2 mutations can cause neonatal diabetes, suggesting GLUT2 may have a role in human insulin secretion. Diabetologia. 2012;55(9):2381-5. doi: 10.1007/s00125-012-2595-0. PubMed PMID: 22660720.
58. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews Genetics. 2012;13(7):484-92. Epub 2012/05/30. doi: 10.1038/nrg3230. PubMed PMID: 22641018.
59. Stoffers DA, Thomas MK, Habener JF. Homeodomain protein IDX-1: a master regulator of pancreas development and insulin gene expression. Trends Endocrinol Metab. 1997;8(4):145-51. PubMed PMID: 18406800.
60. Stoffers DA, Ferrer J, Clarke WL, Habener JF. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. Nat Genet. 1997;17(2):138-9. doi: 10.1038/ng1097-138. PubMed PMID: 9326926.
61. Ahlgren U, Jonsson J, Jonsson L, Simu K, Edlund H. beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. Genes Dev. 1998;12(12):1763-8. PubMed PMID: 9637677; PubMed Central PMCID: PMCPMC316911.

62. Park JH, Stoffers DA, Nicholls RD, Simmons RA. Development of type 2 diabetes following intrauterine growth retardation in rats is associated with progressive epigenetic silencing of Pdx1. J Clin Invest. 2008;118(6):2316-24. doi: 10.1172/JCI33655. PubMed PMID: 18464933; PubMed Central PMCID: PMCPMC2373422.
63. Thompson RF, Fazzari MJ, Niu H, Barzilai N, Simmons RA, Greally JM. Experimental intrauterine growth restriction induces alterations in DNA methylation and gene expression in pancreatic islets of rats. The Journal of biological chemistry. 2010;285(20):15111-8. doi: 10.1074/jbc.M109.095133. PubMed PMID: 20194508; PubMed Central PMCID: PMCPMC2865297.
64. Olsson AH, Volkov P, Bacos K, Dayeh T, Hall E, Nilsson EA, et al. Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. PLoS genetics. 2014;10(11):e1004735. doi: 10.1371/journal.pgen.1004735. PubMed PMID: 25375650; PubMed Central PMCID: PMCPMC4222689.
65. Rosengren AH, Braun M, Mahdi T, Andersson SA, Travers ME, Shigeto M, et al. Reduced insulin exocytosis in human pancreatic beta-cells with gene variants linked to type 2 diabetes. Diabetes. 2012;61(7):1726-33. doi: 10.2337/db11-1516. PubMed PMID: 22492527; PubMed Central PMCID: PMCPMC3379663.
66. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Magi R, Reschen ME, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet. 2015;47(12):1415-25. doi: 10.1038/ng.3437. PubMed PMID: 26551672; PubMed Central PMCID: PMCPMC4666734.
67. Dayeh TA, Olsson AH, Volkov P, Almgren P, Rönn T, Ling C. Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia. 2013;56(5):1036-46. Epub 2013/03/07. doi: 10.1007/s00125-012-2815-7. PubMed PMID: 23462794; PubMed Central PMCID: PMC3622750.
68. Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest. 2007;117(8):2155-63. doi: 10.1172/JCI30706. PubMed PMID: 17671651; PubMed Central PMCID: PMCPMC1934596.
69. Hodson DJ, Mitchell RK, Marselli L, Pullen TJ, Gimeno Brias S, Semplici F, et al. ADCY5 couples glucose to insulin secretion in human islets. Diabetes. 2014;63(9):3009-21. doi: 10.2337/db13-1607. PubMed PMID: 24740569; PubMed Central PMCID: PMCPMC4141364.
70. Allum F, Shao X, Guenard F, Simon MM, Busche S, Caron M, et al. Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. Nat Commun. 2015;6:7211. doi: 10.1038/ncomms8211. PubMed PMID: 26021296; PubMed Central PMCID: PMCPMC4544751.
71. Wang T, Guan W, Lin J, Boutaoui N, Canino G, Luo J, et al. A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. Epigenetics : official journal of the DNA Methylation Society. 2015;10(7):662-9. doi: 10.1080/15592294.2015.1057384. PubMed PMID: 26036609; PubMed Central PMCID: PMCPMC4623491.
72. Ziller MJ, Hansen KD, Meissner A, Aryee MJ. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nat Methods. 2015;12(3):230-2, 1 p following 2. doi: 10.1038/nmeth.3152. PubMed PMID: 25362363; PubMed Central PMCID: PMCPMC4344394.
73. Bell RE, Golan T, Malcov H, Amar D, Salamon A, Liron T, et al. Enhancer methylation dynamics contribute to cancer plasticity and patient mortality. Genome research. 2016. doi: 10.1101/gr.197194.115. PubMed PMID: 26907635.
74. Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schubeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. Nature. 2015;528(7583):575-9. doi: 10.1038/nature16462. PubMed PMID: 26675734.
75. Gao W, Fu Y, Yu C, Wang S, Zhang Y, Zong C, et al. Elevation of NR4A3 expression and its possible role in modulating insulin expression in the pancreatic beta cell. PLoS One. 2014;9(3):e91462. doi: 10.1371/journal.pone.0091462. PubMed PMID: 24638142; PubMed Central PMCID: PMCPMC3956668.
76. Kimura H. Histone modifications for human epigenome analysis. J Hum Genet. 2013;58(7):439-45. doi: 10.1038/jhg.2013.66. PubMed PMID: 23739122.

77. Li B, Carey M, Workman JL. The role of chromatin during transcription. Cell. 2007;128(4):707-19. doi: 10.1016/j.cell.2007.01.015. PubMed PMID: 17320508.
78. Siggens L, Ekwall K. Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. J Intern Med. 2014;276(3):201-14. doi: 10.1111/joim.12231. PubMed PMID: 24605849.
79. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. Nature reviews Genetics. 2011;12(1):7-18. doi: 10.1038/nrg2905. PubMed PMID: 21116306.

Supporting information captions

S1 Figure. Technical validation, unsupervised PCA analysis and luciferase assays. A) Technical validation of WGBS library prep kits and sequencing chemistries. Correlations between WGBS data generated using two different library preparation kits (TruSeq DNA Methylation Kit and NEXTflex™ Bisulfite Library Prep Kit), and using two different sequencing chemistries (Illumina type 3 and type 4 chemistry). **B**) Unsupervised PCA analysis of WGBS data from human pancreatic islets. **C**) The selected genes for functional luciferase assays show altered DNA methylation of promoter DMRs in human pancreatic islets from type 2 diabetic donors. **D-E**) Luciferase assays to functionally study the impact of altered DNA methylation on transcriptional activity. The luciferase assay results showed that methylation of the (**D**) *KIF3A* and (**E**) *TMED6* promoters with SssI and HhaI/HpaII (the number of methylation sites for each enzyme is indicated in the figure) resulted in reduced transcriptional activity. ***$P < 0.001$ as analyzed with an unpaired two-tailed t-test (n=5).

S1 Table. Characteristics for 75 donors of pancreatic islets for biological replication.

S2 Table. Sheet A) Pyrosequencing assay used for biological replication of *PDX1,* **Sheet B)** promoter sequences (2000 bp) used for insertion in the luciferase reporter vector and **Sheet C)** cDNA sequences for overexpression experiments in the 832/13 INS-1 β-cell line.

S3 Table. **Sheet A)** Sequencing information and alignment statistics of WGBS in human pancreatic islets. **Sheet B)** Correlation between DNA methylation of 224,645 CpG sites covered by both WGBS and 450K[1] (forward strand) from the same 14 samples of human pancreatic islets.

S4 Table. Differentially methylated regions (DMRs) in human pancreatic islets from donors with T2D.

S5 Table. T2D associated DMRs in human pancreatic islets annotated to the 65 known T2D candidate loci identified by the Diagram consortium (Morris et al).

S6 Table. Overlap between T2D associated DMRs and genome-wide maps of histone modifications, generated by the Roadmap Epigenomics Consortium (Kundaje et al), and enhancer regions in human pancreatic islets (Pasquali et al). **Sheet A)** H3K4me3 **Sheet B)** H3K9ac **Sheet C)** H3K27ac **Sheet D)** H3K4me1 **Sheet E)** H3K27me3 **Sheet F)** H3K9me3 **Sheet G)** H3K36me3 **Sheet H)** Enhancers.

S7 Table. Enrichment of transcription factor recognition sequences (known/novel motifs) in our T2D associated DMRs based on HOMER.

S8 Table. Overlap between T2D associated DMRs and transcription factor binding sites in human pancreatic islets: **Sheet A)** FOXA2, **Sheet B)** MAFB**, Sheet C)** NKX2.2**, Sheet D)** NKX6.1, **Sheet E)** PDX1

S9 Table. Overlap between T2D associated DMRs in human pancreatic islets and regions shown to have dynamic methylation between human cell types and tissues by Ziller et al.

S10 Table. T2D associated DMRs annotated to genes that also show altered RNA expression in human pancreatic islets from donors with T2D.

S11 Table. KEGG pathways with enrichment of genes with both significant DMRs and differential expression in islets from donors with T2D vs. normoglycemic controls.