

From the Institute of Epidemiology II,
Research Unit Molecular Epidemiology,
Helmholtz Zentrum München, German Research Center for Environmental Health
Head: Prof. Dr. rer. biol. hum. Annette Peters

Epigenetic analysis of type 2 diabetes and measures of glucose metabolism

Thesis

Submitted for a Doctoral Degree in Natural Sciences
at the Faculty of Medicine
Ludwig-Maximilians-University, Munich

by

Jennifer Heidi Kriebel

from Leverkusen, Germany

2015

**Printed with approval from the Faculty of Medicine,
Ludwig-Maximilians-University, Munich**

Supervisor/Examiner: Prof. Dr. rer. nat. Thomas Illig

Co-Examiner: Prof. Dr. rer. nat. Axel Imhof

Dean: Prof. Dr. med. dent. Reinhard Hickel

Date of oral examination: 16.10.2015

Table of contents

Summary	VI
Zusammenfassung	VIII
List of abbreviations	X
1. Introduction	- 1 -
1.1. Pathogenesis and diagnosis of type 2 diabetes.....	- 1 -
1.2. Genetic and environmental contributions to the development of diabetes	- 3 -
1.2.1. Family based linking analysis	- 4 -
1.2.2. Candidate gene studies	- 4 -
1.2.3. Genome-wide association studies	- 5 -
1.2.4. Missing heritability and different approaches to overcome the gap.....	- 6 -
1.3. Epigenetics	- 9 -
1.3.1. Background on epigenetics	- 10 -
1.3.2. Epigenetic mechanisms.....	- 12 -
1.3.2.1. RNA interference and histone modification	- 12 -
1.3.2.2. DNA methylation	- 13 -
1.3.3. Epigenetics and diabetes	- 15 -
1.4. Aim of the study	- 17 -
2. Material and Methods	- 19 -
2.1. Ethics statement.....	- 19 -
2.2. Study population	- 19 -
2.2.1. Nested case-control study	- 19 -
2.2.2. Cross-sectional study	- 21 -
2.2.3. Assessment of glucose and insulin level as well as diabetes status	- 22 -
2.3. Quality control of samples for bisulfite conversion	- 23 -
2.3.1. Agarose gel electrophoresis.....	- 23 -
2.3.2. Determination of DNA concentration	- 24 -
2.3.3. Amelogenin test for sex determination	- 25 -
2.4. Bisulfite conversion.....	- 26 -
2.4.1. Principle.....	- 26 -
2.4.2. Laboratory procedure.....	- 27 -
2.5. Genome-wide DNA Methylation analysis.....	- 29 -
2.5.1. Principle.....	- 29 -
2.5.2. Laboratory procedure for Infinium HumanMethylation450 BeadChip, following Infinium HD Methylation protocol	- 32 -
2.5.3. GenomeStudio	- 36 -
2.6. Replication and Fine mapping	- 36 -
2.6.1. EpiTYPER®	- 37 -
2.6.1.1. Principle.....	- 37 -
2.6.1.2. Laboratory procedure.....	- 39 -
2.6.2. Pyrosequencing	- 42 -
2.6.2.1. Principle.....	- 42 -
2.6.2.2. Laboratory procedure.....	- 44 -
2.7. Statistical analysis	- 45 -

2.7.1.	Preprocessing and QC.....	- 45 -
2.7.1.1.	Genome-wide DNA Methylation analysis.....	- 45 -
2.7.1.2.	Replication and Fine Mapping	- 47 -
2.7.1.3.	Comparison SQN vs BMIQ	- 47 -
2.7.2.	Statistical analysis	- 48 -
2.7.2.1.	Nested case-control study	- 48 -
2.7.2.2.	Replication for LOLIPOP study.....	- 49 -
2.7.2.3.	Cross-sectional study	- 50 -
2.7.2.4.	Gene expression analysis	- 51 -
2.8.	Ingenuity Pathway analysis	- 52 -
2.9.	Material.....	- 53 -
2.9.1.	Laboratory equipment	- 53 -
2.9.2.	Chemicals, reagents, enzymes and assays	- 55 -
2.9.3.	Kits	- 56 -
2.9.4.	Computer software and programs.....	- 57 -
2.9.5.	Online databases and programs	- 57 -
3.	Results	- 58 -
3.1.	Association with incident T2D.....	- 58 -
3.1.1.	Study characteristics	- 58 -
3.1.2.	Genome-wide DNA methylation analysis.....	- 60 -
3.1.3.	Replication and fine mapping.....	- 63 -
3.1.4.	Pathway analysis	- 68 -
3.1.5.	Replication for LOLIPOP study.....	- 69 -
3.2.	Association with measures of glucose metabolism	- 72 -
3.2.1.	Study characteristics	- 72 -
3.2.2.	Genome-wide DNA methylation analysis.....	- 74 -
3.2.3.	DNA methylation quintile analysis.....	- 79 -
3.2.4.	Pathway analysis	- 83 -
3.2.5.	Gene expression analysis	- 85 -
4.	Discussion	- 88 -
4.1.	Association with incident T2D.....	- 88 -
4.1.1.	Main findings.....	- 88 -
4.1.2.	Involvement of <i>AKT2</i> in diabetes.....	- 91 -
4.1.3.	Involvement of remaining associated CpG sites in diabetes	- 93 -
4.1.4.	Accumulation of DNA methylation data in different pathways and link to diabetes	- 94 -
4.1.5.	Involvement of genes found in Indian Asians and Europeans in diabetes ..	- 95 -
4.2.	Association with measures of glucose metabolism	- 96 -
4.2.1.	Main findings.....	- 96 -
4.2.2.	Involvement of <i>ABCG1</i> in diabetes	- 98 -
4.2.3.	Involvement of remaining associated CpG sites in diabetes	- 100 -
4.2.4.	Accumulation of DNA methylation data in different pathways and link to diabetes	- 102 -
4.3.	General challenges.....	- 103 -

4.4.	Strengths and limitations	- 104 -
4.5.	Conclusion	- 105 -
4.6.	Outlook	- 106 -
5.	Literature	XII
6.	Appendix	XXIII
7.	List of Tables	XLV
8.	List of Figures.....	XLVIII
9.	Description of own contribution	L
10.	List of publications and presentations	LI
10.1.	Publications included in this thesis.....	LI
10.2.	Publications not included in this thesis.....	LI
10.3.	Poster presentation with results of this doctoral thesis	LIII
10.4.	Poster presentations prior to doctoral thesis	LIV
11.	Acknowledgement (Danksagung)	LV
12.	Eidesstattliche Erklärung	LVI

Summary

Diabetes mellitus is one of the most widespread diseases worldwide. About 381.8 million people have the disease, and it is estimated that just as many have undetected diabetes or are at high risk (pre-diabetes). In order to better understand the disease and its associated burden it is important to understand the molecular disease mechanisms. The genetic aspects of type 2 diabetes (T2D), which impact the majority of diabetes cases, has been comprehensively studied in genome-wide genetic variation studies. Thus, today we know about a large number of genes involved in disease development. However, to my knowledge, there are no systematic genome-wide studies on epigenetic regulation mechanisms in whole blood and participants with European ancestry. Therefore, the objectives of this thesis are to identify an association between DNA methylation and i) incident T2D, and ii) measures of glucose metabolism in whole blood using samples from the Cooperative Health Research in the Region of Augsburg (KORA) study.

In the first part of this thesis the methylation degrees of six CpG sites were genome-wide significantly associated with incident T2D [annotated to *CASZ1*, *TMEM57*, *VIM*, *C14orf182*, *AKT2*, and one unknown (adjusted p-values= 6.2×10^{-3} to 1.9×10^{-2})]. Three CpG sites flanking the leading CpG site annotated to *AKT2* and one to *C14orf182* were found to be significantly associated with the phenotype in the replication. These associations seem to be driven by body mass index (BMI). Furthermore, performing pathway analysis with DNA methylation, data pathways such as “Nerve Growth Factor Signaling”, “Integrin Signaling” or “G-protein gamma/beta Signaling” were detected, each of which can be linked to diabetes. Using data from KORA for replication of the results from the London Life Science Prospective Population (LOLIPOP) study, which consists of Indian Asians and European individuals, it was possible to replicate five of the seven significant CpG sites (annotated to *TXNIP*, *PHOSPHO1*, *SOCS3*, *SREBF1*, and *ABCG1*). The difference in findings can be explained by differing study designs, especially concerning different ethnicities, matching criteria for cases and controls, and sample sizes. In the second part of this thesis associations between the methylation degrees of 15 CpG sites and measures of glucose metabolism were detected [four

for fasting glucose, seven for fasting insulin, ten for homeostasis model assessment-insulin resistance (HOMA-IR); adjusted p-values = 6.8×10^{-5} to 0.043]. In addition, DNA methylation at cg06500161 (*ABCG1*) was significantly associated with fasting insulin, fasting glucose, 2-hour glucose, and HOMA-IR (adjusted p-values = 8.1×10^{-4} , 6.8×10^{-5} , 1.3×10^{-3} , and 7.5×10^{-5} , respectively). BMI explains around 30% of these associations. Furthermore, a significant association between the CpG sites found for the phenotypes named above and 2-hour insulin was observed in a subset of samples. These were also in part significant after additional adjustment for BMI. Analyzing associations between DNA methylation and gene expression in another subset, CpG site cg06500161 (*ABCG1*) showed an association with *ABCG1* gene expression level (adjusted p-value = 1.1×10^{-9}). Additionally, an accumulation of the top 1,000 CpG sites in pathways such as “Leptin Signaling in Obesity”, “Ephrin A/B Signaling”, “Netrin Signaling”, and “Phospholipase C Signaling”, which can be linked to diabetes, was found.

In conclusion, this thesis indicates an association between DNA methylation, incident T2D/measures of glucose metabolism in whole blood which seems to be BMI mediated. These results can help to better understand the underlying pathogenesis of T2D and measures of glucose metabolism and can thus be used to potentially develop biomarkers and new therapies for T2D.

Zusammenfassung

Diabetes mellitus ist eine der meist verbreiteten Krankheiten weltweit. Etwa 381,8 Millionen Menschen haben diese Krankheit und es wird geschätzt, dass noch einmal genau so viele einen undiagnostizierten Diabetes bzw. ein hohes Risiko für Diabetes (Prädiabetes) haben. Um Diabetes und somit auch dessen Bürde besser zu verstehen ist es wichtig die molekularen Mechanismen der Krankheit zu klären. Der genetische Part von Typ 2 Diabetes (T2D), welcher fast den ganzen Anteil der Diabetiker ausmacht, wurde umfassend durch genomweite genetische Variationsstudien untersucht. Aufgrund dieser ist bisher eine große Anzahl an Genen bekannt, die bei der Entstehung der Krankheit involviert sind. Jedoch wurde bisher, nach meinem Kenntnissstand, von keiner systematischen genomweiten Studien zu epigenetischen Regulationsmechanismen in Vollblut von Europäern berichtet. Daher war das Ziel dieser Arbeit eine Assoziation zwischen der DNA Methylierung und i) inzidenten T2D sowie ii) Maßen des Glukosemetabolismus zu identifizieren. Dies wurde in Vollblutproben von Probanden der Kooperativen Gesundheitsforschung in der Region Augsburg (KORA)-Studie durchgeführt.

Im ersten Projekt dieser Doktorarbeit wurden sechs genomweit signifikant assoziierte CpG Stellen mit inzidenten T2D gefunden [annotiert zu *CASZ1*, *TMEM57*, *VIM*, *C14orf182*, *AKT2* und eine nicht annotierte (adjustierte p-Werte = 6.2×10^{-3} bis 1.9×10^{-2}]. Für drei CpG Stellen, welche die zu replizierende CpG Stelle annotiert zu *AKT2* flankieren, sowie für eine für *C14orf182*, konnte eine signifikante Assoziation mit dem Phänotypen gefunden werden. Diese Assoziationen sind vorwiegend durch den „body mass index“ (BMI) vermittelt. Ebenso konnten durch Pathwayanalysen Signalwege wie „Nerve Growth Factor Signaling“, „Integrin Signaling“ oder „G-protein gamma/beta Signaling“ gefunden, welche mit Diabetes verbunden werden können. In der London Life Science Prospective Population (LOLIPOP), welche aus Indischen Asiaten und Europäern besteht, konnten unter der Verwendung der KORA-Daten fünf der sieben signifikanten CpG Stellen (annotiert zu *TXNIP*, *PHOSPHO1*, *SOCS3*, *SREBF1* und *ABCG1*) repliziert werden. Die unterschiedlichen Ergebnisse könnten aufgrund des unterschiedlichen Studiendesigns, im speziellen die verschiedenen Ethnizitäten, Matchingkriterien für die Fälle und Kontrollen oder die Probandenanzahl, erklärt werden. Im zweiten

Projekt dieser Doktorarbeit wurden für insgesamt 15 CpG Stellen, eine signifikante Assoziation zwischen dem Methylierungsgrad und Maßen des Glukosemetabolismus gefunden [vier für Nüchtern glukose, sieben für Insulin, zehn für „homeostasis model assessment-insulin resistance“ (HOMA-IR), adjustierte p-Werte = 6.8×10^{-5} bis 0.043]. Zusätzlich konnte eine signifikante Assoziation der DNA Methylierung für cg06500161 (*ABCG1*) und Nüchterninsulin, Nüchtern glukose, 2h-Glukose und HOMA-IR aufgezeigt werden (adjustierte p-Werte = 8.1×10^{-4} , 6.8×10^{-5} , 1.3×10^{-3} und 7.5×10^{-5}). Der BMI erklärt ungefähr 30% dieser Assoziationen. Des Weiteren konnte bei der Analyse der CpG Stellen, welche bei den zuvor erwähnten Phänotypen gefunden wurde, in einer Teilgruppe von Probanden eine signifikante Assoziation mit 2h-Insulin aufgezeigt werden. Diese sind zum Teil ebenso nach der Adjustierung für den BMI signifikant. Bei der Analyse der Assoziation zwischen der DNA Methylierung und der Genexpression in einer weiteren Teilgruppe, zeigte cg06500161 (*ABCG1*) eine Assoziation mit dem *ABCG1* Genexpressionslevel (adjustierter p-Wert = 1.1×10^{-9}). Zusätzlich wurde eine Anhäufung der Top 1000 CpG Stellen in Signalwegen wie „Leptin Signaling in Obesity“, „Ephrin A/B Signaling“, „Netrin signaling“ oder „Phospholipase C Signaling“ gezeigt, welche mit Diabetes in Verbindung gebracht werden können.

Schlussfolgernd weisen die Ergebnisse dieser Doktorarbeit auf eine Assoziation von der DNA Methylierung und inzidentem T2D sowie Maßen der Glukosemetabolismus in Vollblut hin, welche zum größten Teil durch den BMI vermittelt ist. Diese Erkenntnisse können bei dem besseren Verständnis der T2D und Maßen des Glukosemetabolismus zugrunde liegenden Pathogenese helfen und daher für die Verwendung von potentiellen Biomarkern sowie der Entwicklung von neuen Therapien für T2D verwendet werden.

List of abbreviations

μl	microliter
APS	adenosine 5' phosphosulfate
ATP	adenosine triphosphate
B-H	Benjamini-Hochberg
BMI	body mass index
BMIQ	beta mixture quantile normalized
bp	base pairs
CEU	Utah Residents with Northern and Western European ancestry
CNV	copy number variation
CpG	cytosine-phosphate-guanine site
CRP	c-reactive protein
Da	dalton
dATP	desoxyriboadenosine triphosphate
DCCT	Diabetes Control and Complications Trial
DIAGRAM	Diabetes Genetics Replication and Meta-analysis Consortium
DNA	deoxyribonucleic acid
dNTP	desoxyribonucleotide triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FIN	Finnish in Finland
GBR	British in England and Scotland
GWAS	genome-wide association studies
h	hour
H ₂ O	water
HbA1c	glycated hemoglobin
HDL	high density lipoprotein
HOMA-IR	homeostasis model assessment-insulin resistance
HPLC	high performance liquid chromatography
IBS	Iberian population in Spain
kb	kilobase
KORA	Cooperative Health Research in the Region of Augsburg
l	liter
LD	linkage disequilibrium
lincRNAs	long intergenic non coding RNAs
lncRNA	long non-coding RNAs
LOLIPOP	London Life Science Prospective Population
MAF	minor allele frequency

MAGIC	Meta-analysis of Glucose- and Insulin-related Traits Consortium
MeDIP	methylated DNA immunoprecipitation
min	minute
miRNAs	micro RNA
ml	milliliter
mM	millimolar
NGSP	National Glycohemoglobin Standardization Program
ng	nanogram
OGTT	oral glucose tolerance test
OR	odds ratio
PCR	polymerase chain reaction
piRNA	PIWI-interaction RNA
pmol	picomole
POC	point-of-care
PPi	pyrophosphate
RCF	relative centrifugal force
RLGS	restriction landmark genomic scanning
RNA	ribonucleic acid
rpm	revolutions per minute
RRBS	reduced representation bisulfate sequencing
RT	room temperature
SAP	shrimp alkaline phosphatase
SD	standard deviation
SE	standard error
sec	seconds
snoRNA	small nuclear RNA
SNP	single nucleotide polymorphism
SQN	subset quantile normalized
T2D	type 2 diabetes
TRIS	tris (hydroxymethyl) aminomethane
TSI	Toscani in Italia
TSS	transcription start site
UTR	untranslated region
UV	ultraviolet
V	volt
WHO	World Health Organization
QC	quality control

1. Introduction

The prevalence of type 2 diabetes (T2D) has increased substantially over the last decades and is expected to increase from 381.8 million in 2013 to 591.9 million people in 2035 worldwide. These numbers include an estimate of 175 million unreported diabetics, mainly T2D patients. T2D poses a heavy burden on health systems. In 2013, 548 billion US dollars were spent on treating diabetes and managing of complications worldwide. It is estimated that this number will exceed 627 billion US dollars in 2035 [1]. In 2013, it was estimated that 56 million people are diabetic in Europe with an overall estimated prevalence of 8.5% [2]. 80% of the people affected by diabetes are living in low- and middle-income countries [1]. The reported numbers underscore the importance of investigating this disease to gain a better understanding of pathogenic mechanisms and to develop new strategies for therapies and prevention.

1.1. Pathogenesis and diagnosis of type 2 diabetes

T2D is a metabolic dysfunction with multifactorial causes of disease. Hallmarks of the pathogenesis are impaired insulin secretion or reduced insulin sensitivity or a combination of both, eventually leading to chronic hyperglycemia. Previously, T2D was also called “non-insulin dependent diabetes” or “adult onset diabetes” [3, 4]. Reduced insulin sensitivity can lead to insulin resistance in insulin responsive tissues (liver, muscle, adipose tissue) [3-5]. Initially, this can be compensated by an increased release of insulin by β -cells. In the later course of the disease, glucose uptake is decreased in insulin sensitive tissues, which finally leads to increased blood glucose levels [4, 6]. Ultimately, this results in hyperinsulinemia, impairment of β -cell function, and eventually manifestation of diabetes (figure 1) [4, 7, 8]. T2D can be predicted by low-grade systemic inflammation [9] and it often remains undiagnosed over years as hyperglycemia develops gradually and typical symptoms are not recognized by patients in the early phase of the disease [3, 4, 10].

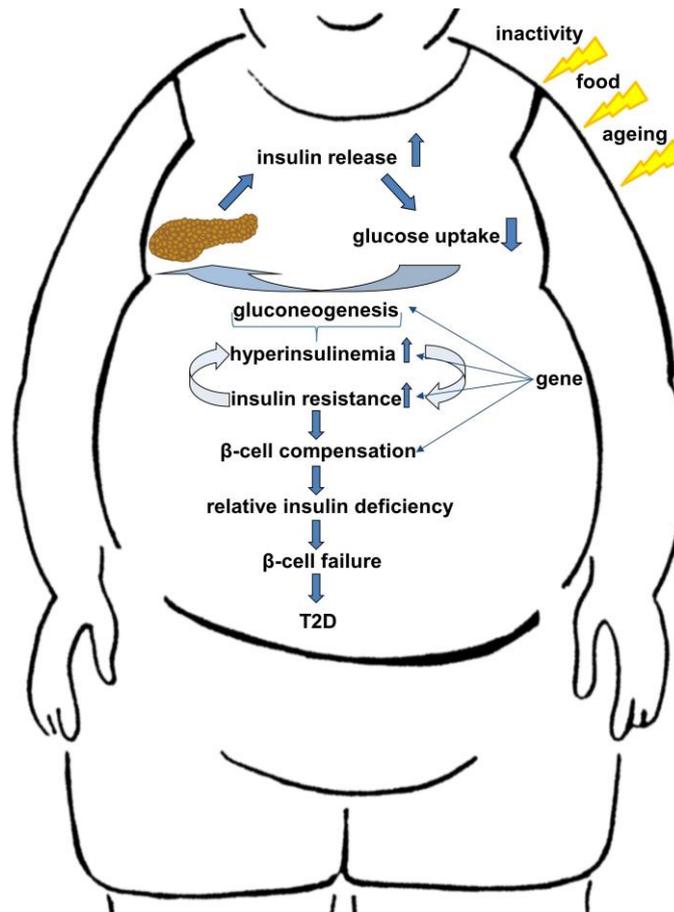


Figure 1: Pathogenesis of T2D. (modified by [11])

At present, the criteria for T2D diagnosis are as follows:

Fasting plasma glucose ≥ 7.0 mmol/l (126 mg/dl) and/or
 2-hour plasma glucose ≥ 11.1 mmol/l (200 mg/dl) and/or
 Random plasma glucose ≥ 11.1 mmol/l (200 mg/dl) [4, 12, 13]

In addition, glycated hemoglobin (HbA1c) has also been recently used as a criterion for diagnosis [4, 12, 14]. After a repeated value of 6.5% T2D can be diagnosed. A value between 5.7-6.4% indicates an increased risk for T2D, but lower values do not fully exclude the disease [4, 12]. The use of HbA1c is controversial, and the World Health Organization (WHO) recommends that a standardized assay be used for quality control [15]. Recently, the American Diabetes Association (ADA) recommends the use of National Glycohemoglobin Standardization Program (NGSP) certified methods or that are identifiable to the Diabetes Control and Complications Trial (DCCT), whereas point-of-care (POC) HbA1c assays should not

be used. Furthermore, they delimit that age, race, and diseases like anemia have to be considered using the HbA1c as a diagnosis criterion for diabetes [12]. It was found that a HbA1c cut of $\geq 6.5\%$ detects one-third fewer undiagnosed people with diabetes than with the fasting glucose cut point [16]. The German Diabetes Association recommends a combination of glucose- and HbA1c values, so as to benefit from both, besides self-administered questionnaires [17].

1.2. Genetic and environmental contributions to the development of diabetes

T2D is influenced by genetic as well as environmental factors (figure 1) [2, 4, 7, 18, 19]. The latter, which is only touched on in this thesis as it is not the main focus, include increased food intake, physical inactivity, aging, and smoking [2, 20]. As obesity is one of the major risk factors for T2D [5, 21], it is plausible that most T2D patients are obese, and this in turn can lead to insulin resistance. Glucose homeostasis in obese persons can be improved by reducing body weight, increasing physical activity, and/or by medical treatment. However normal conditions cannot be re-established [4]. Furthermore, other factors like microbiota composition [22], ethnicity [23] or circadian rhythm [24] can have an influence on the development of T2D.

In addition to environmental factors, genetic factors play an important role in the development of T2D [8, 19, 25]. However, the numbers of T2D patients have been increasing in the last decades, a time window too short for genetic drift, and it is likely that most of the increased prevalence is due to environmental changes [5, 26]. This is in line with the increasing prevalence of diabetes in times of economic change and urbanization, as these in turn lead to lifestyle change.

There are substantial differences in prevalence of T2D between ethnicities. For example it is 2-fold higher for Asians or Pima Indians residing in Western countries compared to Europeans [19]. The gene with the strongest association to T2D in the European population is *TCF7L2*, whereas for the Asian population this is *KCNQ1* [19, 27-30]. Furthermore, genes such as *UBE2E2* and *C2CD4A/4B* have been

identified as risk factors only in the Asian population [31], whereas *THADA* or *MNRT1B* seems to play a more important role in Europeans [30, 32].

The investigation of T2D susceptibility genes can be subdivided in three major historical phases [5, 19, 30], which will be described in detail in the following subchapters.

1.2.1. Family based linking analysis

In family based linkage analysis chromosomal regions associated with T2D are identified by genetic markers in family pedigree. This approach was most useful for the identification of genetic variants for extreme, monogenic forms of early manifestation of the disease [19]. Due to these analyses, new insights into processes responsible for normal glucose homeostasis, energy balance, function of pancreatic β -cells, and hypothalamus were obtained [30]. However, family based linkage analyses are inappropriate for investigating polygenetic diseases. Nevertheless, in this area the high-risk regions could not be consistently associated with T2D in a study population study [19, 32].

1.2.2. Candidate gene studies

The second phase was driven by hypothesis based approaches focusing on candidate genes. Genes with a plausible involvement due to their function were analyzed in case-control association analyses [7, 19]. On one hand these studies were more effective than linkage analysis, despite their lower power due to smaller sample size. On the other hand they were restricted to specific causal variants [30]. Grant *et al.* had shown an association of *TCF7L2* with T2D analyzing 228 microsatellite markers in an Icelandic as well as Danish population [27]. Although more than 100 candidate genes were investigated, only variants in *PPAR γ* , *KCNJ11*, and *TCF7L2* showed consistent and reproducible associations with T2D [7, 8, 19, 27, 30, 33].

1.2.3. Genome-wide association studies

The breakthrough for the detection of T2D susceptibility variants started in 2007 with the era of genome-wide association studies (GWAS). In these studies systematic, large scale analyses are performed to evaluate the association between common single nucleotide polymorphism (SNPs) and diseases [26, 30], thus constituting a hypothesis free approach to detect genetic determinants of disease across the whole genome [19, 26]. Technological advances, such as the development of DNA Chips, enabled a cost-effective approach to analyze many samples genome-wide in a short period of time [7, 19]. Furthermore, analysis of the human genome sequence by the HapMap project, enabled the detection of common SNPs and patterns of linkage disequilibrium (LD) [5, 8, 20, 32]. These features together with the progress in biostatistics were the basis for GWAS [19]. The 1000 Genome project extended the human genome sequence by whole-genome and exon-targeted sequencing [8, 20, 32].

Beside newly identified genes, known T2D genes such as *TCF7L2*, *KCNJ11*, and *PPAR γ* were confirmed by these studies [19]. To date, *TCF7L2* is the gene with the strongest association with T2D in the European population [odds ratio (OR) =1.46 per allele, $p=5.4 \times 10^{-140}$] [34].

Using the local LD patterns from HapMap, new imputation methods have been developed to determine genotypes at untyped SNPs due to directly typed SNPs. About 2.2 million SNPs can be analyzed genome-wide combining different genotyping platforms [8]. A statistical p-value of 5×10^{-8} is adopted for genome-wide significance level, reflecting the standard p-value of 0.05 assuming a Bonferroni correction with 1 million statistical tests [32].

Soon it became obvious that the effect sizes for additional loci were only modest and therefore larger study populations were needed to detect them [8, 32]. For this reason the diabetes genetics replication and meta-analysis consortium (DIAGRAM) were built to perform meta-analysis on T2D and related traits [meta-analysis of glucose- and insulin- related traits consortium (MAGIC)] [8, 26]. Not only meta-analysis for T2D reveal T2D susceptible genes, but also meta-analysis for related traits in healthy individuals identified associated loci [8].

Consistently defined phenotypes across cohorts as well as well-structured, and obligatory analysis plans are elementary for successful meta-analysis [8]. Furthermore, the findings of *KCNQ1* in East-Asians point out that multiethnic studies might help to discover further T2D associated variants [8, 35], which was confirmed by DIAGRAM in 2014 [36].

So far, 88 genetic susceptible loci are identified for T2D mostly by GWAS summarized in figure 2 [33, 36-38]. The majority of these loci are associated with β -cell function, whereas only few are related to insulin sensitivity [5, 32, 39, 40].

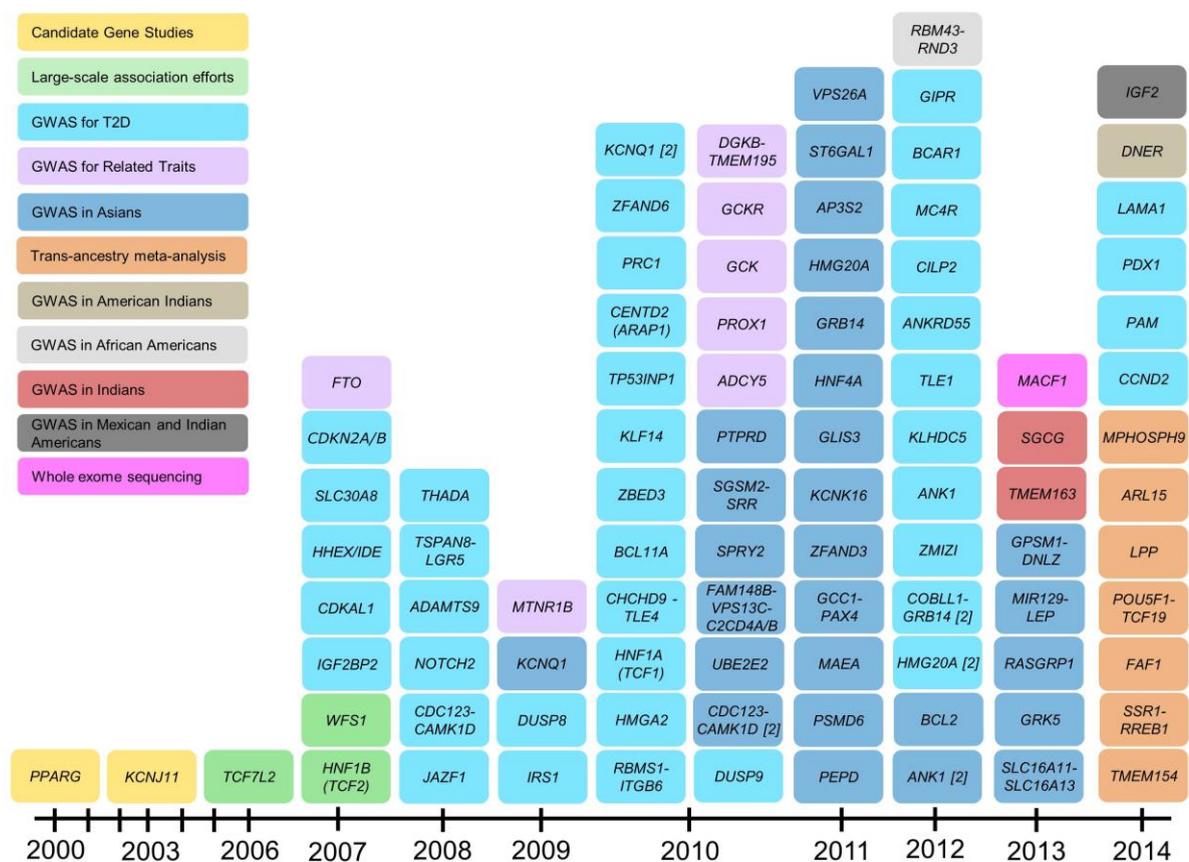


Figure 2: Identified T2D susceptible loci divided into year of discovery. (modified [33, 36-38])

1.2.4. Missing heritability and different approaches to overcome the gap

Heritability estimates for T2D and related traits differ largely because of variety of population or type of study with estimates ranging between 15-85% [5, 41-43]. All 88 genetic susceptible loci identified for T2D identified so far [33, 36-38] can only explain 5-10% of the estimated heritability [25, 32, 39, 44]. All studies on T2D and

related traits of heritability are consistent in reporting a substantial gap between estimated and explained heritability, the so-called “missing heritability” (figure 3) [26, 32]. However, it has become almost impossible to increase the power of GWAS on the basis of sample size and thus, GWAS seem to have reached their limitations in analyzing T2D as well as other diseases. Therefore, it is necessary to develop new approaches.

Groop and colleagues postulate that the missing heritability is overestimated in the present studies based on the observation that it is often derive from twin or family studies. Often the GWAS didn't take family clustering into account [26].

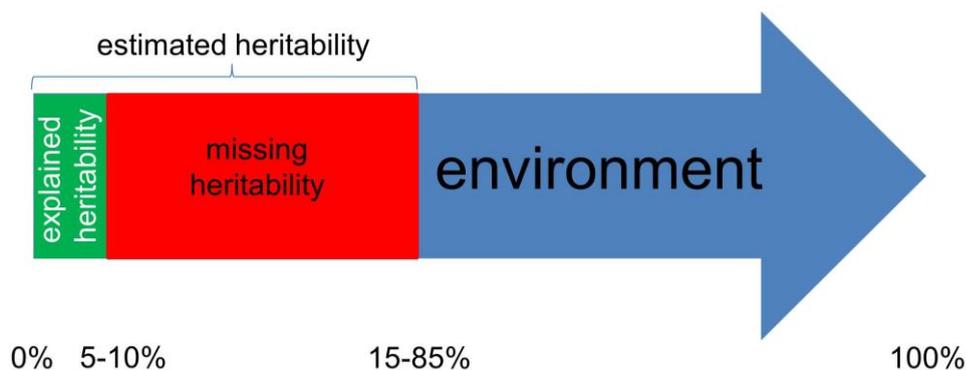


Figure 3: Contribution of genetic and environmental factors to development of T2D and related traits and missing heritability. The number of percentage represents the proportion of genetic and environmental contribution as well as the explained heritability so far.

Potential solutions to elucidate “missing heritability” can be the evaluation of genetic variants in multiethnic studies [25, 32, 35]. Efforts in this direction were recently demonstrated by the DIAGRAM consortium that detected seven novel T2D susceptibility loci by performing a trans-ancestry approach. Further, a study by Saxena *et al.* demonstrated an association between *BCL2* and T2D using a multiethnic approach [36, 45]. Furthermore, large-scale analysis in multiethnic studies could also support the detection of genes with borderline association, i.e. those that could not be found with the commonly used threshold of $p < 5 \times 10^{-8}$ for GWAS [32] as well as low-frequency risk variants (minor allele frequency 0.3-5.0%) with relatively large effects can also help to clarify missing heritability [32, 35, 46].

Furthermore, rare variants can help to explain missing heritability and studies have started to detect these variants using next generation sequencing (for example whole-exome and whole-genome sequencing) [5, 25, 26, 32, 33, 35, 46, 47]. For

example, Albrechtsen *et al.* show an association of *COBLL1* and *MACF1* and T2D performing whole-exome sequencing in Danish individuals [48]. Until today, rare as well as low-frequency variants are not studied systematically in GWAS, mostly because of power limitations due to small sample size [20].

It is also possible that genes do not only influence the risk of diabetes by modifying insulin sensitivity directly, but also by influencing the interaction of a person with the environment [5]. This so called gene-environment interaction may play a role in disease pathogenesis [10, 47]. A different perception of unhealthy diet or the comparison of the microbiome may serve as examples for gene-environment interaction [5]. For the analysis of gene-environment interaction large studies with well-defined phenotypes are necessary [26], to investigate how specific loci, that play a role in the disease, change with exposure to different environmental factors like diet or exercise [10].

Gene-gene interaction is postulated to generate “phantom heritability” and therefore contribute to missing heritability by building a wrong number in the denominator [49]. Gene-gene interaction may be an explanation for the small success of replication of genetic associations of complex diseases [49-52]. The standard statistical analysis, which is generally applied, often didn't take the interaction between different loci into account. It is suggested to analyze significant SNPs against each other or all other SNPs included in the study, but often the sample size is a limitation factor [26].

Furthermore, there is the assumption that genetic variants associated with traits are hidden by epistatic processes, which means that these variants can affect heritability by changing or influencing other processes [25]. Another potential approach to clarify the missing heritability is the analysis of the differences between maternal or paternal transmission of the risk allele (parent-of origin transmission of risk allele). For instance, it was observed that risk alleles of the *KCNQ1* and *KLF14* genes influence the risk of T2D stronger if they are transmitted from the mother compared to the father [53, 54]. On the other hand the paternal allele can have a protective influence and therefore making it impossible to detect in recent case-control studies [26].

Furthermore, additional processes by which a protein can be changed can help to identify the missing heritability. At present GWAS are mainly based on the assumption that genes are transcribed and then translated into the protein [26, 55]. In addition, it is important to build networks, pathways or interaction with metabolites or other omics-data [55-57]. In this context studies on epigenetic modifications, which can result in a modified gene expression [5, 26, 47], alike in a tissue-specific context [25], are promising approaches to elucidate part of the missing heritability.

1.3. Epigenetics

All healthy cells in an organism share the same genotype, however they differ in a broad spectrum of functions and phenotypes. Recent studies point out that besides the DNA sequence other mechanisms are involved in defining phenotypes, without changing the genetic.

This is known as “epigenetics”, which comprises DNA methylation, histone modification and RNA interference (figure 4). Epigenetic mechanisms influence the gene expression and therefore influence the development of diseases [5, 26, 47].

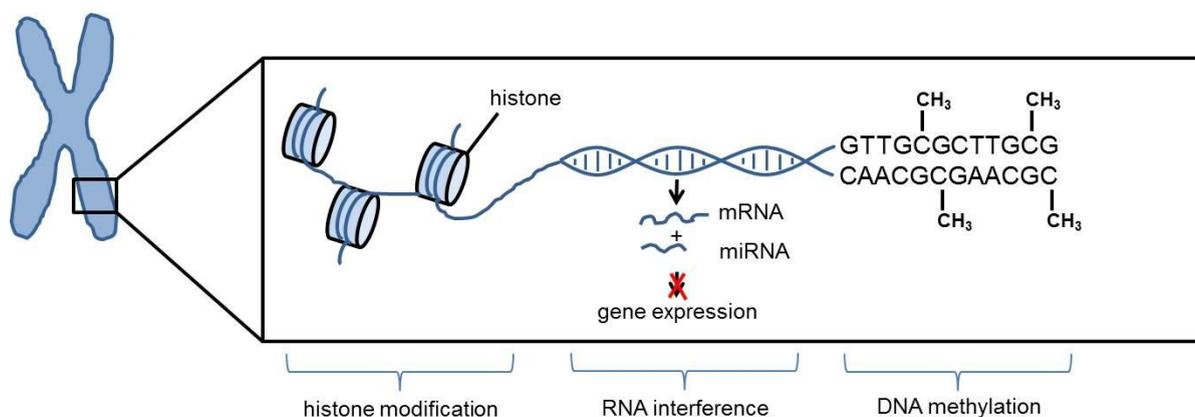


Figure 4: Overview of epigenetic mechanisms. Possible chemical modifications like acetylation or methylation are presented as  at histones (modified after [58]).

1.3.1. Background on epigenetics

The word epigenetic includes the Greek prefix “epi” which can be translated with over, before, upon [59] and can therefore be seen as a second instance of information in addition to genetics. Conrad Waddington, one of the pioneers in this field, realized that genetic and developmental biology are more related than previously accepted and created the word epigenetic in the middle of the 20th century. For him, epigenetic and embryology are not really two different fields [59]. Epigenetics is defined as “The study by mitotically and/or meiotically heritable changes in the gene function that cannot be explained by changes in DNA sequence” [60].

Epigenetic modifications like DNA methylation, histone modification, and non-coding RNAs alter gene expression and can persist for the whole lifespan [20, 26, 61]. Furthermore, it has been assumed that DNA methylation and imprinting are involved in unique parent-of-origin transmission of risk alleles [26, 61].

Epigenetic characteristics differ in some aspects from genetics. For example epigenetic modifications are often tissue specific [25, 59, 62], whereas the same genetic code can be found in every tissue. Furthermore, genetic changes are stable, whereas epigenetic changes are often reversible. Most common environmental factors don't change genetic status, but impact epigenetic status [59].

Epigenetic inheritance is a controversial topic. It can be divided into mitotic inheritance (from one cell generation to another) [63] and meiotic inheritance (inheritance between generations). Furthermore, epigenetic inheritance can be divided into two parts: inter- and transgenerational (figure 5) [63, 64]. In plants it was shown that epigenetic modifications are meiotically inherited [65]. Dunn and colleagues show an influence of maternal high-fat diet on body weight and insulin sensitivity in offspring of the F3 generation [66]. One further example is the agouti mouse. The agouti gene is one of the genes coding for the coat color and various versions of this gene lead to different coat colors. For the agouti viable yellow (*Avy*) gene it was shown that the degree of methylation influences the coat color [63, 67]. The disposition of the offspring phenotype is influenced by the maternal phenotype. This maternal epigenetic effect is more common in the case of incomplete deletion

of the modification than inhibition [67]. In humans inheritance is still a subject of ongoing discussions in the field. Different studies demonstrate an intergenerational effect. For example Heijmans *et al.* show that prenatally exposed to famine individuals from the Dutch Hunger Winter had a reduced DNA methylation of the imprinted *IGF2* gene compared to their siblings not exposed to famine [68]. Furthermore, it was shown a higher neonatal adiposity in individuals where the grandmother was malnourished during pregnancy [69, 70]. There is evidence that epigenetic changes induced by the environment can be inherited transgenerationally [61]. In their review, Heard and Martienssen discussed different studies and aspects dealing with inheritance in plants, animals, and humans. They conclude that there is plenty of evidence of transgenerational epigenetic inheritance in plants and animals like nematodes, but for humans there is still only little support. The discussion also concerned papers where it is shown that it has a different influence if the female or male line is exposed to an environmental factor [64].

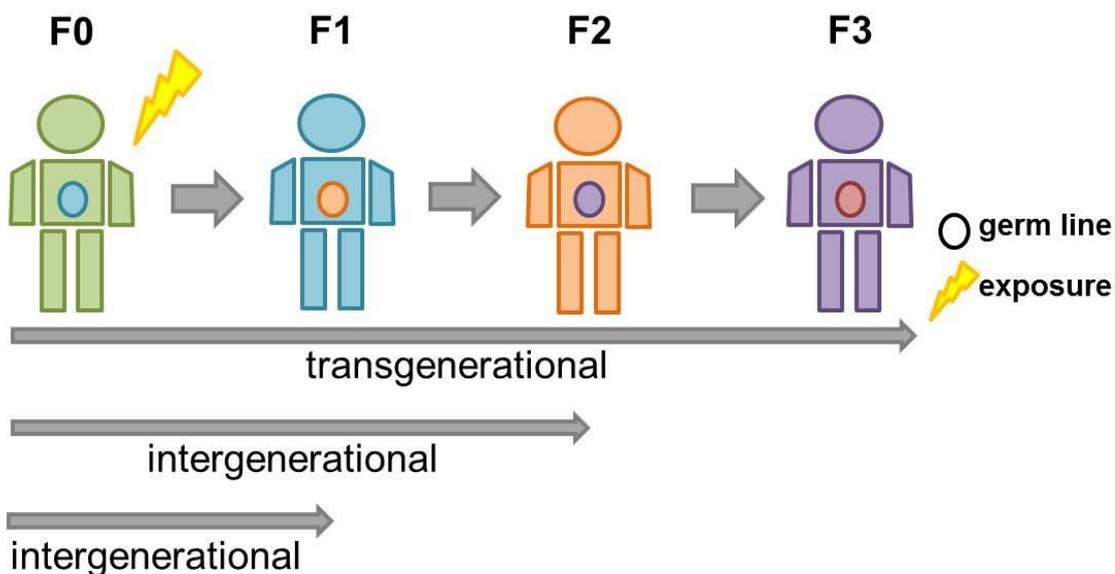


Figure 5: Overview of different inheritance possibilities. The different colors of germ lines represent the cells which are passed on to the next generation.

1.3.2. Epigenetic mechanisms

1.3.2.1. RNA interference and histone modification

RNA interference

In RNA interference non-coding, single stranded RNA fragments like miRNAs bind mRNAs at the “RNA induced silencing complex”, and thereby inhibit the translation and influence gene expression [20, 26, 71]. Binding of miRNAs to transcripts depend on the sequence, intra-molecular structure as well as SNPs that can prevent binding [26]. This mechanism can be observed during cell proliferation, apoptosis or tumor metastasis. Furthermore, non-coding RNAs can modify chromatin structure and therefore genome stability [71]. Recently, miRNAs have been of interest as potential therapeutic target due to the involvement of miRNAs in human disease development. Furthermore, other non-coding RNAs like piRNAs (PIWI-interacting RNAs), snoRNAs (small nucleolar RNAs), lincRNAs (long intergenic non coding RNAs), and lncRNA (long non-coding RNAs) influence the gene expression and the development of diseases [26].

Histone modification

The kind of packing of chromatin, consisting of histone-DNA complex, influences DNA transcription, replication, and repair mechanisms [71]. Histones can be modified by acetylation, methylation, ubiquitination, phosphorylation, glycosylation, sumoylation, and adenosine diphosphate ribosylation leading to changes of chromatin structure and thereby gene expression [63, 71]. Most histone acetylation results in a more open chromatin structure, whereas deacetylation is generally associated with a compacted chromatin structure and therefore transcriptional repression [72]. But no generalization can be done, as it depends on the position and which histone or amino acid is modified [63, 73].

1.3.2.2. DNA methylation

At the moment, the most prominent and best studied epigenetic mechanisms is DNA methylation. A broad range of well-established methods for analyzing it is commercial available. Due to this and the fact that DNA methylation is the focus of this thesis, this epigenetic modification is described in more detail below.

Background

In 1969, Griffith and Mahler postulate that DNA methylation is an important foundation for long-term memory in the brain [74]. In 1975, Riggs as well as Holliday and Push hypothesized that DNA methylation has strong effects on gene expression. Changes in gene expression were explained by switch-on/switch-off mechanisms of genes during development [75, 76].

DNA methylation is involved in different key processes like gene imprinting, embryonic development, X-chromosome gene silencing, genome stability, and regulation of gene expression [59, 77-82]. In past studies changes in DNA methylation in promotor cytosine phosphorylated guanosine (CpG) sites was under extensive investigation, but now the focus is shifting to CpG island shores [71].

CpG sites are often clustered in so called CpG-islands and are located in promoter regions, 1st exons of genes or in upstream CpG island shores (less than 2 kb distance to CpG islands) or upstream shelves (more than 2 kb distance to CpG islands) [83, 84]. CpG islands can be found within the 5`promoter region of around 60% of expressed genes [62]. DNA methylation modifies the regulation and expression of genes and thus the stability of the genome [59, 63, 71]. Gene promoter hypermethylation is mostly associated with reduced gene expression [85, 86]. But there are studies published showing that this dogma is not generally adaptive. Recent studies also show that DNA methylation within shores is also associated with reduced gene expression [87, 88], whereas methylation in the gene body can lead to increased gene transcription [89, 90].

Mechanisms

DNA methylation mainly occurs at the cytosine, where a methyl group binds to the 5`carbon of the cytosine in CpG dinucleotides, resulting in 5`-methylcytosine [71].

This process is catalyzed by DNA-methyltransferases, binding methylgroups donated from S-adenosyl methionine [77] and can be reversed by DNA methylases [91]. Three functional forms of DNA-methyltransferases (DNMT1, DNMT3A, DNMT3B) are shown to be involved in de novo DNA methylation [92]. Additionally, it was observed that ten-eleven translocation proteins can oxidize 5`-methylcytosine to 5`-hydroxymethylcytosine and further to 5`-formylcytosine and subsequently to 5`-carboxycytosine [93-96]. Also, it was demonstrated that 5-hydroxymethylcytosine and 5`-formylcytosine are implicated in Alzheimer`s disease or Huntington`s disease [93, 97]. 5`-methylcytosine tend to spontaneously hydrolytic deaminate to thymine [98, 99] and therefore it seems obvious that 5`-methylcytosine is underrepresented. DNA methylation is also known in non-CG context, showing enrichment in gene bodies, and it is suggested that around one-quarter of all methylation in embryonic stem cells are linked to this phenomena [100].

The degree of DNA methylation can be affected by genetic as well as environmental factors [71, 101] and therefore impact the development of diseases. Figure 6 illustrates how genetic as well as epigenetic factors might influence the development of diabetes and how the genetic can influence the epigenetic status and thereby gene expression.

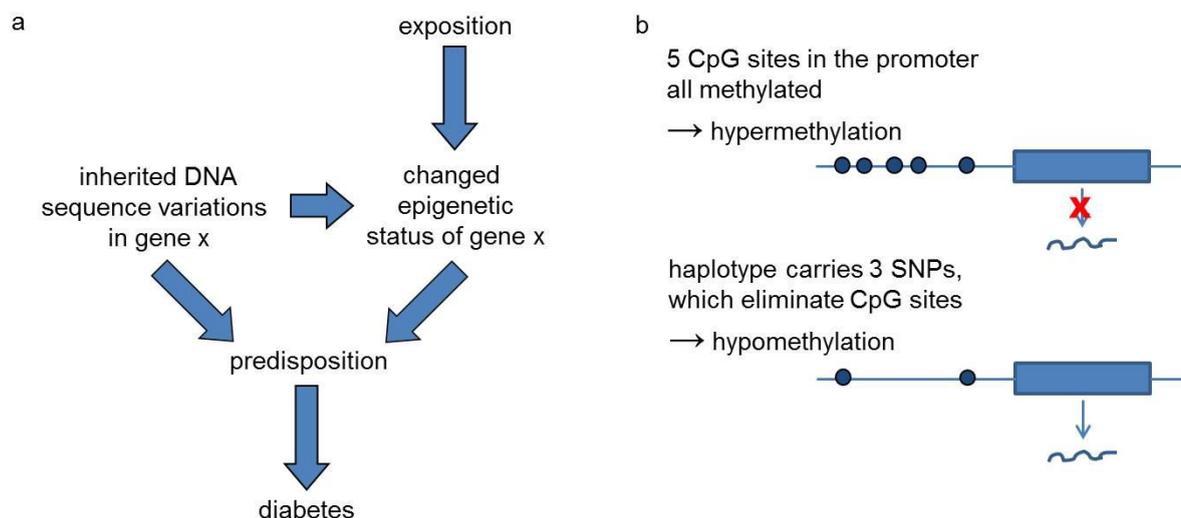


Figure 6: Interaction of genetic and epigenetic factors. A) Influence of genetic and epigenetic factors on the development of diabetes. B) Interaction of genetic and epigenetic status on gene expression.

Many environmental factors contribute to diseases by changes in DNA methylation [71]. Also, it was shown that smoking influences the degree of DNA methylation [102]. Furthermore, it was demonstrated that prenatal maternal stress induced DNA methylation in offspring's [103] or that exposure with air pollutions like black carbon and sulfate were associated with methylation pattern in the asthma pathway [104]. Further examples in context of diabetes are described in chapter 1.3.3.

1.3.3. Epigenetics and diabetes

Small born offspring's due to nutrient manipulation have an increased risk to develop diabetes or obesity in later life. It is assumed that these can be explained by epigenetic changes [105]. Furthermore, human studies suggest that epigenetic mechanisms can explain the influence of intrauterine nutrition and birth weight for later diabetes, obesity, and metabolic syndrome [106]. Because of this observation it might be that heritability of T2D is more likely due to these modifications resulting from maternal environmental exposure as for inherited DNA sequence changes [5]. Different non coding RNA classes like snoRNAs and piRNAs are shown to influence the development of T2D [26]. In cell culture experiments it was shown that glucose transporter-4, which is important for the glucose uptake in adipose tissue and muscle, seems be increased expressed during adipocyte differentiation through demethylation as well as *PPAR γ* gene region [107, 108].

A number of human studies give first evidence for associations between DNA methylation and T2D. Kuroda *et al.* demonstrate in mouse and human studies using pancreatic β -cells a uniquely demethylation of the mouse *Ins2* and human *INS* promotor and insulin promotor-driven reporter gene activity inhibited by methylation of these CpG sites [109]. Barrès *et al.* showed a hypermethylation of *PGC-1 α* in skeletal muscle from T2D patients comparing 17 individuals with normal glucose tolerance, 8 with impaired glucose tolerance, and 17 with T2D [110]. Ling *et al.* could detect a two-fold increased DNA methylation of *PPARGC1A* gene promoter comparing pancreatic islets of ten T2D patients and nine controls that is associated with a reduced *PPARGC1A* mRNA expression [111]. Also, an intra-pair variation in DNA methylation was increased in repetitive regions compared to gene promotor

regions. An increased variation of *LINE1* sequence methylation was associated with different BMIs and 2-hour glucose, whereas there was no association of promoter methylation and phenotype changes. Also methylation changes in promoters of genes related to T2D were observed including *PPARGC1A* in muscle and *HNF4A* in adipose tissue [112]. Furthermore, increased DNA methylation of the *FTO* obesity susceptibility haplotype was found in human whole blood comparing 30 diabetic females and 30 females without diabetes [113] as well as an enrichment of differentially methylated sites in genomic regions that are associated with T2D comparing 710 T2D cases and 459 controls from different studies [114]. Yang *et al.* demonstrates that DNA methylation at ten CpG sites in the distal *PDX-1* promoter and enhancer regions are significantly increased comparing islets from nine T2D patients and 55 non-diabetics. DNA methylation of this gene is negatively correlated with gene expression. Additionally, they show that hyperglycemia decreases gene expression and increased DNA methylation of *PDX-1* as HbA1c correlates negatively with mRNA expression and positively with DNA methylation [115]. Furthermore, Dayeh and colleagues found that regions near to the transcription start site (TSS) are low methylated and regions further away of TSS show higher degree of methylation, comparing the methylome-wide pattern of pancreatic islets from 15 T2D patients and 34 non-diabetics. They also demonstrate that CpG islands were hypomethylated, whereas shelves and open regions are hypermethylated. They found 853 differentially methylated genes, including *TCF7L2*, *FTO*, *KCNQ1*, and 102 of them (including *CDKN1A*, *PDE7B*, *SEPT9*) were differentially expressed [116]. Hall *et al.* compared the methylome of pancreatic islets from 13 donors treated with palmitate vs non-treated cells as well as gene expression and detect that the treatment influence the DNA methylation level over the whole genome and 290 genes showing an altered gene expression had also a changing in DNA methylation level, including *TCF7L2* and *GLIS3* [117]. Furthermore, it was found that 19 of 40 known T2D associated SNPs introduce or remove CpG sites analyzing pancreatic islets from 84 donors and all CpG-SNPs are associated with different DNA methylation of the CpG-SNP site. CpG-SNPs of *TCF7L2*, *KCNQ1*, *CDKN2A*, *ADCY5*, *WFS1*, and *HMGGA2* are also associated with DNA methylation at surrounding CpG sites [118]. Furthermore, Volkmar *et al.* compared in their study pancreatic islets obtained from five male T2D patients and eleven male non-

diabetics and detected significantly differential DNA methylation of 276 CpG sites annotated to promoters of 254 genes. These findings could not be found in whole blood samples. For some of these CpG sites changes in gene expression were additionally observed. In functional annotation of differentially methylated genes and RNAi experiments point out pathways involved in β -cell survival and function [119]. Heijmans *et al.* demonstrated that people who lived during the Dutch hunger winter (1944/1945), compared to their same-gender siblings, who were not exposed to dearth, show reduced DNA methylation in the *IGF2* gene region after 60 years and were therefore coined [68]. Furthermore, it was shown that alterations in DNA methylation patterns of the muscle are dependent on a positive family history with T2D after a six month sport program [120].

1.4. Aim of the study

Determination of epigenetic patterns can potentially identify new genes that are involved in the pathogenesis of T2D and therefore highlight approaches for prevention strategies and treatment. This includes for example the identification of potential targets for medical treatment or the identification of biomarkers for early diagnosis or risk determination.

Recent literature indicates an involvement of DNA methylation in the development of diabetes or influencing related traits. DNA methylation studies on whole blood in a population-based setting analyzing the association with incident T2D or measures of glucose metabolism are lacking at the moment. Therefore two studies were conducted:

1. The first aim of the present thesis was to generate a methylome-wide profile in whole blood using the Infinium HumanMethylation450 BeadChip to identify loci where the methylation degree is associated with incident T2D. Thus, potential biomarkers can be found which can be used to predict an increased risk for T2D. Furthermore, an accumulation of the relevant loci in pathways underlying the pathogenesis of T2D should be found to functionally integrate the results. Additionally, these methylation data were included in the

replication stage of the LOLIPOP study, which investigates the association of DNA methylation and incident T2D in Indian Asians as well as Europeans.

2. The second aim was to determine loci whose methylation degree is associated with measures of glucose metabolism, including fasting glucose, 2-hour glucose from an oral glucose tolerance test (OGTT), HbA1c, fasting insulin, 2-hour insulin, and HOMA-IR (homeostasis model assessment-insulin resistance) and therefore epigenetic markers which are involved in regulation of glucose metabolism and thus T2D. Furthermore, an involvement of the methylation degree of these loci on gene regulation should be detected. Additionally, for the functional integration of the found genes an accumulation of these in pathways underlying the pathogenesis of T2D should be able to be found.

2. Material and Methods

2.1. Ethics statement

The study has been conducted according to the principles expressed in the Declaration of Helsinki. Written informed consent has been given for all participants. The study, including the protocol for subject recruitment and assessment and the informed consent for participants, was reviewed and approved by the local ethical committee (Bayerische Landesärztekammer).

2.2. Study population

The Cooperative Health Research in the Region of Augsburg (KORA) study comprises of a series of independent population-based epidemiological surveys and follow-up studies of individuals living in the region of Augsburg, Southern Germany and has been described in detail elsewhere [121]. No evidence for population stratification in the KORA study was found [122].

2.2.1. Nested case-control study

The participants in the discovery study with incident T2D were selected in a nested case-control design from the baseline studies S3 (1994/95) and S4 (1999-2001) and for replication K12 (1984/85), S2 (1989/90) in combination with data from corresponding follow-up studies F3 (2004/05), F4 (2006-2008), follow up from K12 (1987/88), and S2. All studies were followed up 2009. Cases were defined as participants without diabetes (no self-reported diabetes and no diabetes treatment) at the time-point of the baseline study, but with self-reported diabetes at the follow-up. Self-reports were validated by contacting the physician or additionally with oral glucose tolerance test (OGTT) in S4/F4. In contrast controls were without diabetes at both time-points (figure 7). For each case, one control was drawn randomly using 1:1 matching with replacement [123, 124], stratified for age \pm 2 years, sex, survey, and observation time till diagnosis of diabetes. Some of the study participants were

selected both as a case as well as control, in case a participant who develops diabetes at an earlier time point.

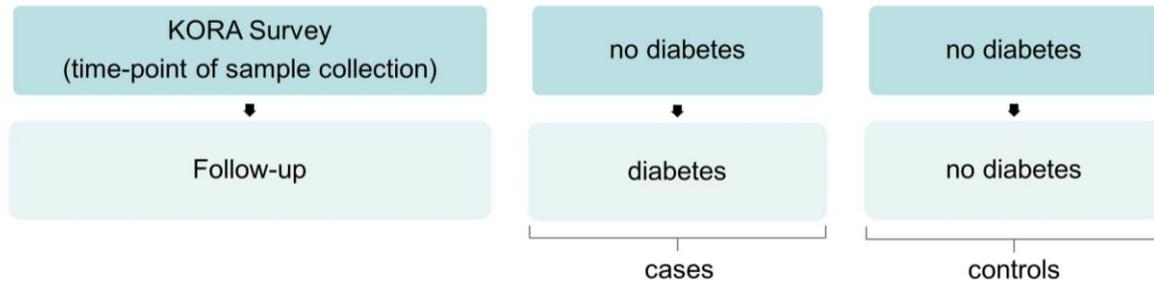


Figure 7: Definition of cases and controls for nested case-control study.

Isolation of DNA was performed with whole blood samples from the baseline studies, which ensured that the samples had no diabetes at the time-point of DNA methylation measurement. Other exclusion criteria were besides no DNA material or DNA degradation, DNA concentration < 30 ng/μl.

From 4,856 participants in S3 and 4,261, 4,022, and 4,940 in S4, K12 and S2, respectively, 330 and 172, 450, 447 had incident T2D, respectively. At the end, 400 participants (200 cases and matched controls; duplicates for control were included) for the discovery study and 802 (401 cases and matched controls, duplicates for controls are included) for the replication study were available for the methylome-wide association as well as fine-mapping analysis for the laboratory process (figure 8). The methylation data of the discovery study were also included in the replication stage of the LOLIPOP study.

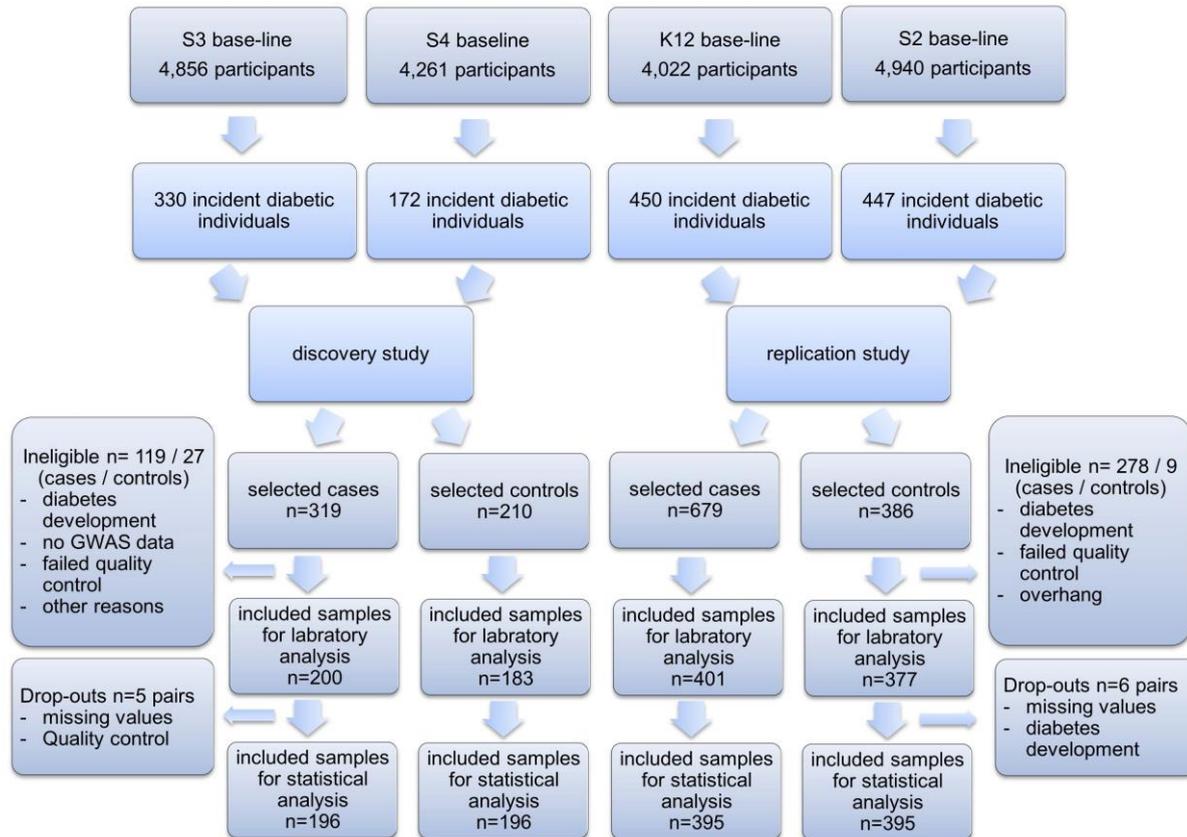


Figure 8: Flow diagram for study population of nested case-control study. The number of controls included for laboratory analysis is slightly smaller than for the statistical analysis, as controls were drawn with replacement. Furthermore, some participants are selected as a case as well as a control, where another case developed T2D at an earlier date.

2.2.2. Cross-sectional study

DNA methylation data for 1,814 participants in the KORA follow-up study by S4 (2006-2008), comprising of 3,080 participants aged between 32 and 81, were randomly selected to generate a DNA-methylation profile [102]. Fifteen samples were excluded due to failed quality control resulting in 1,799 participants available for this study.

Patients with diabetes were excluded based on at least one of the following criteria:

- non fasting (less than 8 hours)
- diabetic patients (fasting glucose ≥ 7 mmol/l and/or 2-hour glucose ≥ 11.1 mmol/l and/or HbA1c $\geq 6.5\%$ and/or undetected/known diabetes and/or diabetes treatment)

- unclear or unknown diabetes status
- missing values in at least one of the outcome parameters and/or covariates
- hs-C-reactive protein > 10 mg/l (as an indicator for an acute infection)
- participants with fasting insulin ≥ 55.718 $\mu\text{IU/ml}$ (corresponding to the 99th percentile) as they clearly represented outliers in the dataset for this variable

The present study includes data from 1,448 non-diabetic individuals with available methylation data for fasting glucose, 2-hour glucose, fasting insulin, and HbA1c as well as 617 and 1,440 for 2-hour insulin and HOMA-IR, respectively (figure 9). Gene expression analyses were conducted on 533 subjects, where gene expression data as well as methylation data were available.

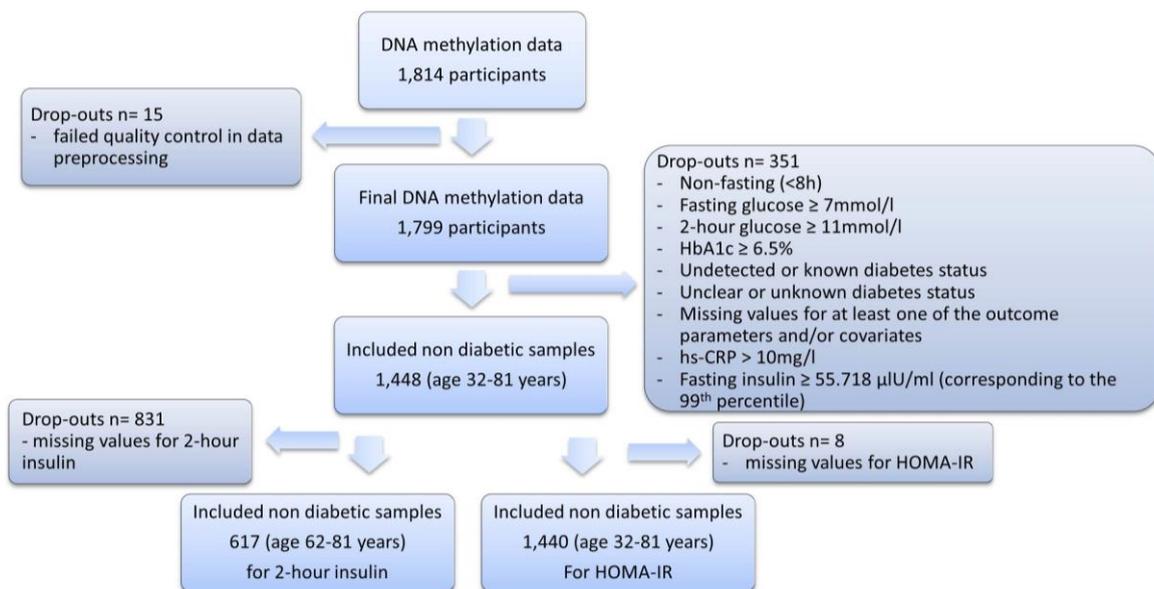


Figure 9: Flow diagram for study population of cross-sectional study.

2.2.3. Assessment of glucose and insulin level as well as diabetes status

Fasting venous blood as well as blood samples after OGTT were collected with minimal stasis, centrifuged, refrigerated at 4 to 8°C, and analyzed in the central laboratory within a maximum of 6 h. Fasting and 2-hour glucose was measured using a hexokinase method (Glucoquant) [125-127]. For glucose the interassay coefficient of variant was 2.5% at 98 mg/dl and 2.1% at 235 mg/dl [125, 126]. HbA1c values in blood were determined using high performance liquid

chromatography (HPLC) (Menarini HA-8160). The coefficient of variation was 1.2% for lower and higher values [126]. Fasting insulin was measured using amikroparticle enzyme immunoassay. HOMA-IR was calculated [fasting plasma glucose (mmol/l) x fasting serum insulin (mU/l) / 22.5] [127].

The OGTT was performed among all non-diabetic participants during the morning hours (07.00-11.00). They were asked to fast for at least 10 h, have no heavy physical activity on the day before, and not to smoke before or during the test. Participants with self-reported diabetes, diabetes treatment, or acute illnesses were excluded. Venous blood samples were taken during the fasting state and afterward 75 g anhydrous glucose was given (Dextro OGT) [126].

Diabetes status was determined in all surveys via self-declaration. Self-reports were validated by contacting the physician or additionally performing the OGTT in S4/F4 [126].

2.3. Quality control of samples for bisulfite conversion

To ensure a good quality for the bisulfite conversion and the following process, DNA quality was detected after the selection of the samples via agarose gel electrophoresis, DNA concentration determination as well as amelogenin polymerase chain reaction (PCR). Samples were used for upstream analysis wherever all quality controls worked well.

2.3.1. Agarose gel electrophoresis

Agarose gel electrophoresis was performed, because long storage time, incorrect storage or many freezing/thawing cycles can lead to DNA degradation. In this method DNA fragments were separated by their size and therefore, the results can be used to determine fragmentation and therefore evaluate the quality of DNA. The size of these fragments can be determined by comparing them with commercial available standards with known fragment sizes. The separation of DNA fragments is based on the migration of nucleic acids to the anode. That effect is due to the

natural negative charge carried on their phosphodiester backbone. Based on the properties of the agarose gel shorter fragments move faster than longer ones.

Determination of the degree of DNA sample degradation was performed with 0.8% agarose gel solved in 1x tris-borate-EDTA (TBE) buffer in the microwave containing 2 µl/100 ml Midori Green, which intercalates with the DNA and can be visualized under UV light.

10x TBE buffer: 108 g Tris
 55 g Boric acid
 9.3 g EDTA pH 8.0
 H₂O_{bidest} up to 1 l

8 µl 6x Loading Dye were added to 2 µl of DNA to sediment them in the gel-well and loaded on the agarose gel. Additionally, 2 µl Lambda DNA/EcoRI + HindIII Marker 3 or GeneRuler 1 kb DNA ladder were loaded. The DNA fragments were separated under a constant electrical field of 60 V for 120 min. After separation, the DNA fragments were visualized and photographed using the Felix 2000. Samples with high degradation were excluded.

2.3.2. Determination of DNA concentration

For determination of DNA concentration the Nanodrop system was used, which is based on spectrophotometry measuring reflection or transmission of material. DNA has a maximum absorption of UV light at 260 nm, whereas proteins have their maximum at 280 nm. Therefore, protein contamination in DNA samples can be estimated due to the A₂₆₀ nm/A₂₈₀ nm ratio. For values between 1.7 and 2.0 protein contamination is excluded to the greatest possible extent and with values > 1.5 salts and phenol contamination. DNA concentration and potential contamination were measured with the 8-channel Nanodrop system following manufacturer's instruction.

2.3.3. Amelogenin test for sex determination

The gender was determined using a PCR amplifying the amelogenin gene, which encodes for a gene responsible for biomineralization during tooth enamel development [128]. Two different forms exist on X and Y-chromosomes, so that women have two copies on the X-chromosome and therefore, one fragment appears on the agarose gel, while men have one copy on the X and Y-chromosome each and therefore, two fragments with different sizes appear on the gel.

Primer	Length of fragment
5'-CTGATGGTTGGCCTCAAGCCTGTG-3'	X = 977 bp
5'-TAAAGAGATTCATTAAGTTGACTG-3'	Y = 788 bp

Pipetting scheme for master mix for one 96-well plate (20 µl reaction)

Puffer	200 µl
dNTP (25 mM)	16 µl
Primer rev (100 pmol)	5 µl
Primer for (100 pmol)	5 µl
MgCl ₂	120 µl
Taq-Polymerase	20 µl
H ₂ O	1534 µl

19 µl PCR master mix were added to 1 µl DNA (1:30 diluted), vortexed, and shortly centrifuged. The following PCR program was used:

Table 1: Cycling protocol for amelogenin PCR.

Amelogenin PCR		
Temperature	Duration	Number of cycles
95°C	15 min	1
95°C	30 sec	40
62°C	30 sec	
72°C	1 min	
72°C	10 min	
20°C	1 min	1

Afterwards the PCR product was loaded on a 1.5% agarose gel, diluted at 1 x TBE puffer, to check the bands and furthermore the gender. The gel runs at 90 V for 90 min (for more details see also chapter 2.3.1.).

2.4. Bisulfite conversion

For bisulfite conversion EZ-96 DNA MethylationTM Kit (Shallow-Well Format) was used provided by Zymo Research.

2.4.1. Principle

Bisulfite treatment was developed by Shapiro and Hayatsu 1970 and it is based on the fact that bisulfites attack cytosines rapidly, but 5-methylcytosine is attacked slowly [129]. After treating denatured single-stranded DNA with bisulfite, unmethylated cytosines will be converted into uracil and afterwards into thymine during whole genome amplification. Methylated cytosines are protected against the changes. Thus C/T variations were produced based on methylation state and can be determined with methylation analysis tools (figure 10). Chemical basis of bisulfite conversion including sulphonation, hydrolytic deamination, and desulphonation is shown in figure 11.

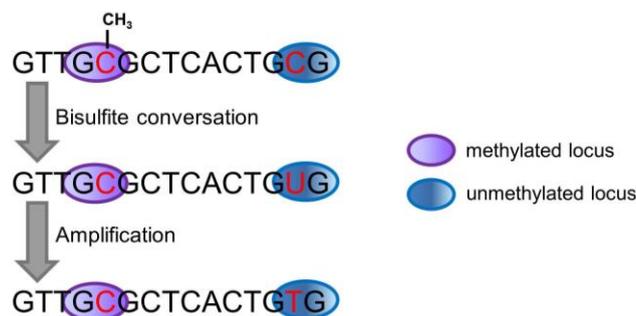


Figure 10: Principle of bisulfite conversion.

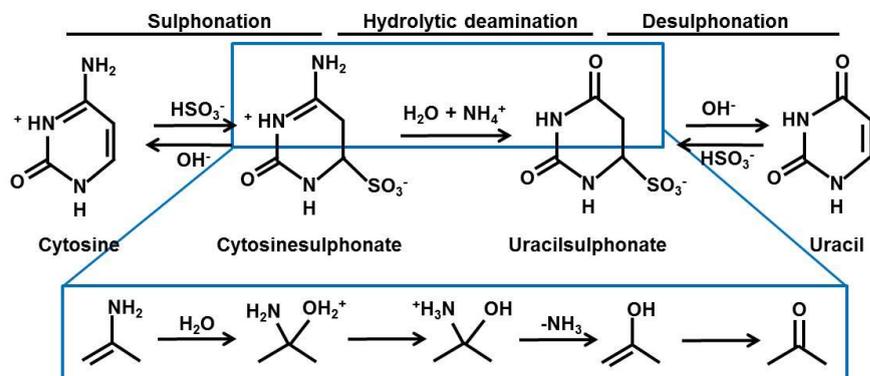


Figure 11: Chemical basis of conversion of cytosine to uracil. The blue box presents the hydrolytic deamination in detail.

2.4.2. Laboratory procedure

As the results for upstream analysis, where 750 ng DNA was used, were comparable to 1,000 ng DNA used for the cross-sectional study (appendix figure 1) for nested case-control study, lower quantities of DNA were used to save DNA material.

Initially, bisulfite conversion reagent was diluted in 7.5 ml water and 2.1 ml dilution buffer and vortexed thoroughly for 10 min at room temperature (RT) till it was completely solved. As the reagent is light sensitive all the steps up till clean-up should be performed in the dark. Furthermore, 144 ml 100% ethanol were added to the wash buffer concentrate.

For bisulfite conversion, 750 ng and 1,000 ng DNA for nested case-control and cross-sectional study respectively were suspended in 45 µl water. 5 µl dilution buffer were added and mixed by pipetting up and down. For upstream EpiTYPER[®] analysis the amount of DNA used depended on the number of investigated regions. The conversion plate was incubated at 37°C for 15 min and afterwards centrifuged. 100 µl of conversion reagent were added to each sample, mixed, centrifuged, and incubated at 50°C for 16-20 h for the following genome-wide DNA methylation analysis using Infinium HumanMethylation450 BeadChip. For analysis via EpiTYPER[®], samples were incubated for 5.5 h. After 1 h at 50°C the samples were heated up to 95°C for 30 sec to ensure that the DNA is single stranded. The correct duration of bisulfite conversion reaction time is important due to different reasons: Bisulfite conversion efficiency depends on denaturation of DNA, as the reaction only takes place with single stranded DNA. Therefore, complete denaturation before and during bisulfite conversion is essential to avoid false-positive signals. On the other hand, overtreatment with bisulfite can lead to degradation of DNA resulting in elevated rates of methylated cytosine to thymine conversion and therefore, possible underestimating DNA methylation. Additionally, it is important to keep the reaction time as brief as possible as it leads to DNA fragmentation and therefore the longer the reaction duration the more fragmentation is likely. Even if larger fragments are needed for the EpiTYPER[®] analysis the reaction time is much shorter compared to Infinium HumanMethylation450 BeadChip.

Table 2: Cycling protocol for Infinium HumanMethylation450 BeadChip respectively EpiTYPER[®] process.

Infinium HumanMethylation450 BeadChip			EpiTYPER [®]	
Temperature	Duration	Number of cycles	Duration	Number of cycles
95°C	30 sec	16	30 sec	20
50°C	1 h		15 min	
4°C	∞		∞	

After (overnight) incubation the samples were immediately incubated for 10 min on ice to stop bisulfite conversion. 400 µl binding buffer were pipetted in the binding plate placed on the collection plate. The converted samples were loaded completely in the well and mixed by pipetting up and down till no smears remained visible. Afterwards the plates were centrifuged for 3 min at 4,500 RCF ($\geq 3,000 \times g$), turned around 180°C, and centrifuged again for 3 min, to ensure that the wells of the binding plate are dry. The flow-through was discarded, 500 µl wash buffer were added and centrifuged again as described before. 200 µl desulphonation buffer were pipetted to the samples and incubated for 15-20 min at RT. Afterwards the plate was again centrifuged as described above, 500 µl wash buffer were added, centrifuged, and another 500 µl wash buffer were added. Samples were centrifuged now for 6 min as described previously. The binding plate was placed onto an elution plate and 15 µl elution buffer were added directly on the filter without touching it. In the subsequent analysis using Infinium HumanMethylation450 BeadChip, elution took place using elution buffer, whereas for EpiTYPER[®] analysis water was used for elution depending on the amount of DNA used (for example for 500 ng inserted DNA, 45 µl water for elution would be used to obtain a final concentration of around 10 ng/µl. Elution was performed in two steps). The samples were incubated for 3 min at RT and afterwards centrifuged for 3 min and finally an additional 2 min after turning the plate. Water as well as elution buffer were warmed to 50°C in the heat oven during the desulphonation to ensure an optimum elution rate as DNA solves better in warm solutions. Converted DNA should be used immediately for upstream analysis, but it can also be stored on a short term basis at -20°C. Converted DNA should not get older than 6 months.

2.5. Genome-wide DNA Methylation analysis

2.5.1. Principle

With regards to the research question, different methods are available to study DNA methylation. These can mainly be divided into genome-wide and fine-mapping analysis. The basis of nearly all of them is bisulfite conversion, which generates a C/T SNP depending on the methylation status of the CpG site (see chapter 2.4.1.). The methods can also be separated into locus-specific analysis (e.g. EpiTYPER[®]), gel-based analysis (e.g. restriction landmark genomic scanning), array-based analysis (e.g. methylated DNA immunoprecipitation), and next generation sequencing-based analysis (e.g. reduced representation bisulfite sequencing) [63, 130].

Genome-wide DNA methylation analyses allow a hypothesis free approach and can be carried out using the Infinium HumanMethylation450 BeadChip, for example. These findings can be then fine-mapped using e.g. EpiTYPER[®] or pyrosequencing. Both methods are described more in detail in chapters 2.6.1.1. and 2.6.2.1.

In the presented study, the Infinium HumanMethylation450 BeadChip was used, as a well-established method for generating genome-wide DNA methylation profiles in high throughput analyses. Furthermore, it is cost effective in comparison to whole-genome bisulfite sequencing for example and allows a hypothesis free approach as well. The Infinium HumanMethylation450 BeadChip enables the determination of the methylation degree of more than 485,000 CpG sites subdivided into 482,421 CpG sites, 3,091 non-CpG sites and 65 SNPs, covering 99% of genes in Reference Sequence. 17 CpG sites per gene were detected on average [84]. Probes were distributed over the whole gene including, promoter region, gene body, 3`UTR, and intergenic region. The promoter region, where around 41% of CpG sites are located, can be subdivided in TSS200 (meaning that the probe is 200 upstream from TSS), TSS1500 (200 till 1,500 bases upstream from TSS), 5`UTR, and 1st exon. Around 31% of the probes are located in CpG island, whereas the remaining CpG sites are located in shores (2 kb flanking CpG island), shelves (2 kb flanking shores) or "others", meaning island independent [83, 84]. 96% of overall CpG sites are covered by the Infinium HumanMethylation450 BeadChip.

Determination of the methylation status in the included samples is based on genotyping the C/T SNP as a result of bisulfite conversion. A single Infinium HumanMethylation450 BeadChip enables the generation of a methylome-wide profile from up to 12 samples in parallel using less than 1,000 ng DNA for bisulfite conversion in the automated Infinium Assay.

Bisulfite converted DNA was whole-genome amplified, fragmented, precipitated, resuspended, and hybridized on the BeadChip. Afterwards the chip was stained and imaged using iScan. One chip needs around 1 h to scan. An overview of the whole process is presented in figure 12.

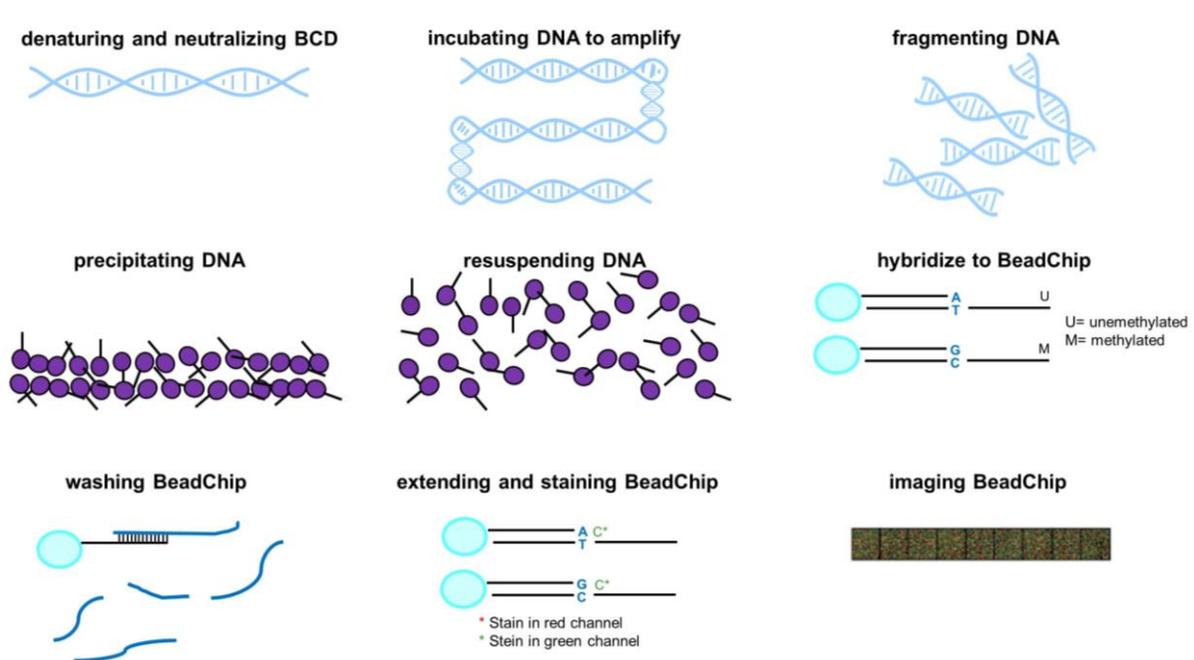


Figure 12: Overview of Illumina laboratory process. BCD: bisulfite converted DNA (adapted after Illumina HD Methylation Assay Guide)

During the hybridization step, the DNA molecules anneal to locus-specific oligomers linked to specific bead types. To measure the methylation at CpG sites two different chemistry technologies were used at the BeadChip namely Infinium I and Infinium II to enable the detection of the huge amount of CpG sites compared to the precursor version, which detected only around 27,000 CpG sites (figure 13).

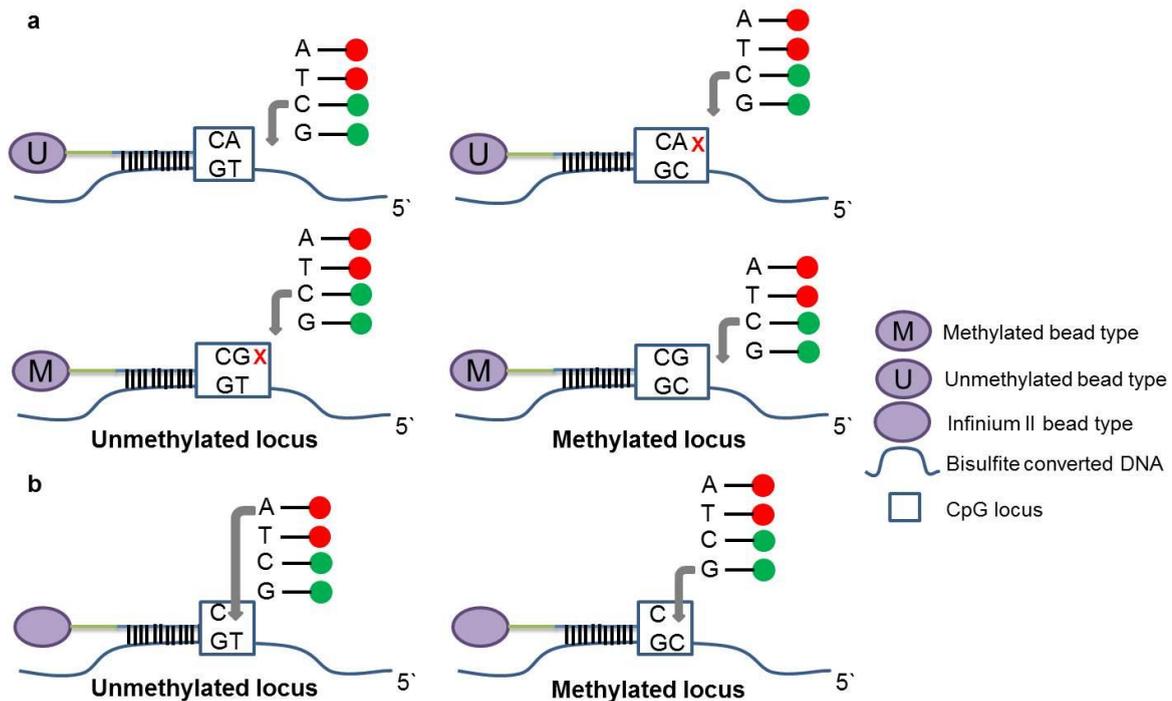


Figure 13: Schematic presentation of the two different chemistry technologies used at the Infinium HumanMethylation450 BeadChip. a) Infinium I assay design including one bead type per CpG locus for unmethylated (thymine) and methylated (cytosine) status each, using allele-specific primer extension. b) Infinium II assay design including only one bead type for determination of both methylated as well as unmethylated status by a single base extension step (modified after [84]).

Infinium I uses two bead types per CpG site, one for the methylated status detecting a cytosine and one for the unmethylated locus detecting a thymine. Allele-specific primer binding takes place followed by a single based extension step with dinitrophenyl labeled nucleotides for adenine/thymine and biotin labeled nucleotides for cytosine/guanosine. The methylated as well as unmethylated bead type for the same CpG locus will incorporate the same type of labeled nucleotide. This will be determined by the base before the interrogated “C” in the CpG locus, and thus will be detected in the same color channel [84]. Signals measured with Infinium I are more stable and have a smaller between-sample variation of β -values than Infinium II. In addition, Infinium II signals are less accurate and reproducible [131]. Infinium II uses only one bead type with a unique type of probe, detecting the methylated as well as the unmethylated locus. The status of CpG locus will be determined directly by a single based extension step. Due to the different labeled nucleotides the signals for the methylated and the unmethylated locus will be measured in the green or red channel [84]. The degree of methylation of a given CpG site will be specified

as a β -value, a continuous number between 0 and 1 calculated as the ratio of methylated probe intensity to overall intensity.

2.5.2. Laboratory procedure for Infinium HumanMethylation450 BeadChip, following Infinium HD Methylation protocol

Day 1 and 2

Generating bisulfite converted DNA

Bisulfite converted DNA was generated as described in chapter 2.4.2. converting unmethylated cytosine over uracil to thymine, creating the C/T SNP, which was measured during the following process.

Day 2

Amplification of bisulfite converted DNA

MSA1 (multi sample amplification 1 mix), RPM (random primer mix) and MSM (multiple sample amplification master mix) were completely thawed at RT. The hybridization oven was preheated at 37°C.

Starting with 4 μ l bisulfite converted DNA in a half-deep well plate, 20 μ l MSA1 were added as well as 4 μ l 0.1 N NaOH. Afterwards the plate was closed with a cap mat, vortexed for 1 min with 1,600 rpm, and centrifuged at 280 x g for 1 min at RT. Samples were incubated for 10 min at RT. This step is time crucial to prevent DNA degradation by NaOH. 68 μ l RPM and 75 μ l MSM was added, closed with a cap mat, and mixed 10 times by inverting. The plate was centrifuged at 280 x g for 1 min, and incubated for 20-24 h at 37°C in the hybridization oven.

Day 3

Fragmentation

FMS (fragmentation solution) was completely thawed at RT and the heat block was preheated at 37°C.

The plate was removed from the hybridization oven and centrifuged at 280 x g for 1 min. 50 µl FMS were added to the samples and closed with a cap mat. The plate was vortexed at 1,600 rpm for 1 min and centrifuged at 280 x g for 1 min, followed by incubation for 1 h in the heat block at 37°C.

Precipitation with isopropanol

PM1 (precipitation solution) was brought to RT.

100 µl PM1 were added to the plates followed by a vortexing step at 1,600 rpm for 1 min. Substances of PM1 intercalate with the DNA and enables a visualization of the pellet after precipitation. Samples were incubated for 5 min at 37°C and centrifuged afterwards at 280 x g. 300 µl isopropanol were added to the plate. It was closed with a new cap mat and mixed by inverting 10 times. Afterwards the plate was incubated for 30 min at 4°C, centrifuged at 3,000 x g for 30 min at 4°C. The supernatant was decanted (after checking for blue pellet) by quickly inverting the plate. It was tapped several times for 1 min on an absorbent pad until all wells were completely drained. The plate was dried upside down on a tube rack for 1 h at RT.

Resuspension

RA1 (resuspension, hybridization, and wash solution) was thawed completely (solution has to re-dissolved). The hybridization oven was preheated at 48°C and the heat sealer was switched on. The centrifuge was cooled down to 4°C.

46 µl RA1 was added to the plate and sealed with an aluminum seal using a heat sealer and incubated for 1 h at 48°C in the hybridization oven, followed by vortexing at 1,800 rpm for 1 min and centrifugation at 280 x g for 1 min.

Hybridization

The heat block was preheated at 95°C. 12 resuspended samples can be hybridized on one BeadChip and therefore up to 8 BeadChips are used for one plate.

Samples were denatured for 20 min at 95°C using a heat block, to enable the annealing of single-stranded samples to locus-specific 50mers, which are linked to one of the around 485,000 CpG sites detectable with the Infinium

HumanMethylation450 BeadChip. During this incubation step the Hyb Chambers were prepared. BeadChip Hyb Chamber gaskets were placed into the Hyb Chambers. 400 µl PB2 (humidifying buffer used during hybridization) were dispensed into the humidifying buffer reservoirs in the Hyb Chamber. Right away the lid was placed on the Hyb Chambers to prevent evaporation. After 20 min incubation the plate was cooled down to RT in 30 min and centrifuged at 280 x g. BeadChips were removed from the zipblock bags and mylar packages and were placed on the Robot BeadChip Alignments Fixtures with the barcode end aligned to the ridges on the fixture. At the robot PC MSA4 tasks | hyb-multi BC2 was selected and the right kind and number of BeadChips were chosen. Robot BeadChip Alignments Fixtures with Robot Tip Alignment Guide were placed onto the robot as well as the sample-plate. 15 µl sample volume were pipetted automatically on the BeadChips. Afterwards the BeadChips were removed from the robot and placed in a Hyb Chamber insert, orienting the barcode at the end (it has to match the barcode symbol to the Hyb Chamber insert). Hyb Chamber inserts were placed together with the BeadChip inside the Hyb Chambers. The closed Hyb chambers were incubated for 16-20 h at 48°C in the hybridization oven using the rocker.

To avoid biases matched cases and controls for nested case-control study were placed next to each other on the BeadChip and therefore, on the same bisulfite conversion plate.

Day 4

Washing, Extension and Staining of the BeadChip

XC4 (Xstain BeadChip solution 4) was prepared by adding 330 ml 100% ethanol. XC1, XC2 (Xstain BeadChip solution 1 and 2), STM (superior two-color master mix), ATM (anti-stain two-color master mix) and TEM (two-color extension master mix) were thawed completely. Formamide was prepared by mixing 1.2 ml water, 23.75 ml formamide, and 50 µl EDTA).

XC1, XC2, STM, ATM, 25 ml formamide, 30 ml RA1, and 145 ml XC3 (Xstain BeadChip solution 3) were loaded at the robot. Last three were placed in a bowl. The program Xstain Task was selected and the thermo block was preheated at

32°C. 150 ml PB1 (reagent used to prepare BeadChip for hybridization) were filled in two washing dishes with washing racks. The cover seal was removed from each BeadChip and they were immediately placed in the washing rack submerged with PB1. Afterwards the full washing rack was moved up and down for 1 min. It was moved into the second washing dish with clean PB1 and the washing step was repeated.

BeadChips were loaded on the multi-sample BeadChip alignment fixture with 150 ml PB1 and black frames. The barcode has to align its barcode with the ridges stamped onto the alignment fixture. Afterwards a clear spacer was placed on top of each BeadChip and the alignment bar was placed onto the alignment fixture. A clean glass back plate was put on top of the clear spacer on each BeadChip. The plate reservoir should be at the barcode end of the BeadChip. Metal clamps were attached to the flow-through chambers and end of clear spacers were trimmed using scissors. Final flow-through chambers were loaded on the robot. When the thermo block reached 44°C, the robot was started using the program Xstain Tasks | Xstain HD Bead Chip.

A plastic shell was filled with PB1 and BeadChips (without metal clamps, glass back plate, clear spacer and black frames) were placed on a staining rack. BeadChips were moved up and down 10 times and afterwards soaked for 5 min. XC4 was filled in a second plastic shell and the washing rack was put into it. It was moved up and down 10 times and soaked for 5 min. The staining rack was removed in a smooth, rapid motion and placed directly on a tube rack with the barcodes faced up. Each chip was placed on a rack with the barcode facing up, which was placed in a vacuum desiccator applying vacuum pressure. BeadChips were dried for around 1 h and afterwards the bottom side was cleaned with ethanol to remove any XC4 excess.

Image BeadChip

BeadChips were scanned using the Illumina iScan, which is a two-channel high-resolution laser imager, scanning BeadChips at two wavelengths creating an image file for each channel. Afterwards the intensity values for each bead type were determined via GenomeScan software and a data file was created.

2.5.3. GenomeStudio

GenomeStudio Methylation Module was used to analyze the methylome-wide data from the data files obtained by the BeadArray Reader in combination with the Infinium HumanMethylation450 manifestfile, including information of annotation of CpG sites. Furthermore, samples were first quality checked and methylation data exported for subsequent statistical analysis.

In the nested case-control study GenomeStudio (version 2011.1) with methylation module (1.9.0) was used to process the raw image data. For the cross-sectional study GenomeStudio (version 2010.3) with methylation module (version 1.8.5) was used. Initial quality control of assay performance was undertaken using the “Control Dashboard” provided by GenomeStudio Software, including the assessment of staining, extension, hybridization, target removal, bisulfite conversion I and II, specificity, and negative as well as non-polymorphic control (appendix figure 2). If samples failed the quality control step, the Infinium HD Methylation protocol was repeated. If they failed this quality control step again, they were excluded from downstream analysis.

β -values, a continuous variable between 0 and 1 representing the methylation degree, were exported and used for statistical analysis since methylation in this study is considered to be the independent variable.

2.6. Replication and Fine mapping

To confirm and replicate results of methylome-wide significant associations in the nested case-control study another two methods were used for this project, namely EpiTYPER[®] and pyrosequencing. With both methods a fine-mapping of selected regions can be performed. *TCF7L2* and *CDKAL1* were included as candidate genes. Pyrosequencing was used as a complementary method as it was not possible to design a primer for CpG sites annotated to *CASZ1* and *TMEM57* due to technical reasons.

2.6.1. EpiTYPER[®]

2.6.1.1. Principle

Besides EpiTYPER[®] analysis other fine-mapping methods are available which can be used, such as pyrosequencing or methyl light methyl sensitive PCR. However, EpiTYPER[®] also allows a data analysis in a high through put manner as it is performed in a 384-well format whereas pyrosequencing, for examples, is only available in a 24- or 96-well format. The EpiTYPER[®] uses base-specific enzymatic cleavage in combination with Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS). This speed and accuracy is used in the MassARRAY[®] to determine methylation status and to identify differentially methylated sites through this quantitative analysis. Two base-specific cleavage reactions (T reverse and C reverse) can be selected to generate specific products. During the C cleavage reaction methylated regions are cleaved at every cytosine to create fragments containing at least one CpG site. T cleavage reaction, which was selected in the presented study, cleaved at every thymine at both methylated as well as non-methylated regions. The resulting products from methylated and non-methylated sites have the same length and differ only in their nucleotide composition.

The method is also based on determination of C/T SNPs created by bisulfite conversion, followed by PCR amplification with T7-promoter tagged primers of selected regions. C/T SNPs of the bisulfite treated DNA appears as G/A changes. Afterwards, unincorporated dNTP leftovers from amplification will be neutralized using shrimp alkaline phosphatase (SAP). Then in vitro transcription a base-specific enzymatic cleavage on the reverse strand is performed. The resulting fragments differ in size and mass depending on the sequence changes generated by bisulfite treatment. The methylated and non-methylated status of CpG sites is determined by MALDI-TOF MS and EpiTYPER[®] software (figure 14).

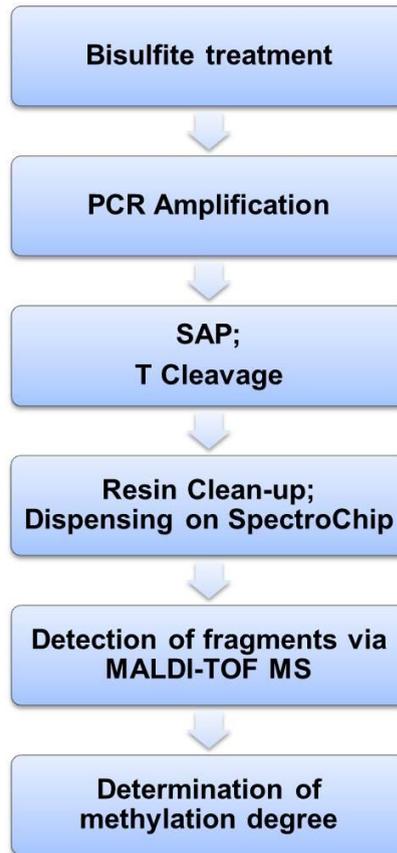


Figure 14: Overview of EpiTYPER[®] workflow.

In MALDI-TOF MS, a laser irradiates the samples on the microchip. Ionized biomolecules are accelerated in an electric field and enter the flight tube. Different ions are separated according to their mass to charge ratio value. Therefore, smaller molecules are faster than larger molecules. Due to the time of flight the mass of every analyt is calculated and will be translated into an allele. Mass difference for non-methylated (adenosine) against methylated (guanosine) fragments for one CpG site is 16 Da.

Three different kinds of mass signal patterns due to methylation can be created:

- methylation generates new cleavage sites resulting in shorter fragments
- methylation generates a replacement of cleavage site into non-cleavage site resulting in a longer fragment
- methylation generates a sequence change in an existing cleavage fragment that doesn't change the cleavage product but results in a mass shift [132].

The SpectroCHIP[®] matrix contains 3-hydroxypicolinic acid, which absorbs the laser energy and generates the ionization of the samples. The matrix inhibits photolytic damage to the fragments, interaction between them or with the sample carrier [133].

Multiple CpG sites in an amplicon of up to 600 bp and changes down to 5% in DNA methylation can be measured. The usage of a 384 microplate format enables a cost-effective high-throughput measurement.

2.6.1.2. Laboratory procedure

Primer design

To design primers for the selected regions EpiDesigner (beta) was used. The following default settings were selected:

Primer temp:	56/62/64 (min/opt/max)
Primer size:	20/25/30 (min/opt/max)
Product size:	100/300/400 (min/opt/max)
Product CpGs:	4
Primer non-CpG C`s:	4
Primer Poly X:	5
Primer Poly T:	8
Selected strand:	both
Mass window:	1500/7000 (low/high)
Analyze CpGs in T reaction:	select
Analyze CpGs in C reaction:	deselect

Sequence of interest (CpG site \pm 300 bp) was passed into EpiDesigner and primer with T7 promoter tags were selected for EpiTYPER[®] analysis. This is carried out according to the number of detected CpG sites at the amplicon and if the CpG sites to be replicated are covered (list of primer- and target sequence attached in appendix table 1 and 2, ordered by Metabion).

Gradient PCR was performed with available bisulfite converted DNA to determine the optimum PCR temperature and therefore to ensure an optimum amplification of regions of interest. These PCR temperatures were used for the following PCR amplification.

Bisulfite treatment

Bisulfite converted DNA was generated as described in chapter 2.4.2. 500 ng genomic DNA was used for bisulfite conversion and at the end diluted with 45 μ l water to get an approximately concentration of 10 ng/ μ l. Matched cases and controls were placed next to each other on the 384 microplate format to avoid biases.

Polymerase Chain Reaction (PCR)

For PCR amplification a master mix with following components was prepared (per sample and amplicon).

H ₂ O	1.42 μ l
10x PCR buffer	0.50 μ l
dNTP mix (25 mM each)	0.04 μ l
Polymerase (5 U/ μ l)	0.04 μ l
Primer Mix (1 μ M reverse and forward)	2.00 μ l

1 μ l bisulfite converted DNA was pipetted on a 384 microplate format and 4 μ l PCR master mix added. Then it was amplified using the cycling protocol presented in table 3.

Table 3: Cycling protocol for PCR amplification for EpiTYPER[®] process.

PCR amplification		
Temperature	Duration	Number of cycles
94°C	4 min	1
94°C	20 sec	
56°C*	30 sec	45
72°C	1 min	
72°C	3 min	1
4°C	∞	

* may be adapted based on amplicon temperature

PCR performance was controlled for each sample via gel electrophoresis using a 3% agarose gel solved in 1xTBE. 0.5 μ l volume of selected sample and 4.5 μ l loading dye were loaded on the gel and run for 30 min at 120 V. PCR product can be stored at -20°C.

SAP Reaction and Transcription Cleavage

Unincorporated dNTP leftovers from PCR amplification were neutralized using SAP enzyme. For each sample, 1.7 μ l RNase free water and 0.3 μ l SAP enzyme were added to the PCR product and incubated for 20 min at 37°C and 5 min at 85°C using a thermo cycler (at the end 4°C forever).

Next, in vitro RNA transcription on the reverse strand was performed, followed by T cleavage transcription to generate fragmented RNA molecules. As this reaction was carried out with RNA molecules the cleavage in fact occurred at a U (uracile). The master mix includes the following components per sample:

RNase free water	3.15 μ l
5 x T7 polymerase buffer	0.89 μ l
T cleavage Mix	0.24 μ l
DTT (100 mM)	0.22 μ l
T7 RNA/DNA polymerase	0.44 μ l
RNase A	0.06 μ l

5 μ l master mix and 2 μ l SAP product were mixed in a new 384 microplate and incubated for 3 h at 37°C using a thermo cycler (at the end 4°C forever). T cleavage product can be stored at -20°C.

Resin Addition, Chip Dispensing and Detection

Finally, clean resin was performed to catch ions in the solution, which could disturb measurement of MALDI-TOF MS. Therefore, T cleavage product was centrifuged and 20 μ l nanopure H₂O pipetted into each well. Furthermore, 6 mg CleanResin was added to the samples. Plates were sealed and twisted for \geq 30 min. Afterwards 20 to 25 nl of the samples were dispenses onto the SpectroChip[®] and additionally a 4-point calibrant was put on position F0, containing four oligonucleotides with known masses. This calibrant includes standardized spectra mass signals from 1,479.0, 3,004.0, 5,044.4, and 8,486.6 Da, which should not differ more than two to three Daltons. Afterwards the chips were loaded in the vacuum lock of the MassARRAY mass spectrometer and ran on a MassARRAY Workstation compact with MassCLEAVE settings. Prior to measuring samples the calibrant was manually controlled.

Analysis of methylation degree

EpiTYPER[®] software was used for editing plates (before MALDI-TOF MS process) and analyzing them. Mass spectra generated by the MassARRAY mass spectrometer were processed and analyzed by the software performing baseline correction, peak identification, and quality assessments. Samples were quality checked using the added controls on the plate. In this project, a negative control (water control), a positive control (bisulfite converted DNA which works already in another analysis) as well as calibration controls (25%, 75% and 100% methylated DNA) were added two times per amplicon on the plate. Furthermore, as the samples were distributed over more than one plate, the performance and reproducibility of each plate was checked by adding plate effect controls (one unique sample) three to four times on the plate. Additionally, one unique sample was placed at the end of each bisulfite conversion plate and analyzed at the end with the last amplicon to control potential differences between bisulfite conversion plates. Samples which failed quality control were repeated and excluded if the replication was not successful. After successful quality control methylation values were exported for statistical analysis.

2.6.2. Pyrosequencing

2.6.2.1. Principle

Pyrosequencing is based on the principle of sequencing by synthesis and provides quantitative data, for example for methylation analysis, by determining released pyrophosphate (PPi) during amplification. Intensity of PPi is equivalent to the amount of incorporated nucleotide. Methylation status of selected regions was determined using pyrosequencing bases of C/T variation generated during bisulfite conversion, using the method described above.

Region of interest is amplified and the strand serving as the template for pyrosequencing will be biotinylated. After denaturation this biotinylated single-stranded DNA is isolated (immobilized on streptavidin-coated beads) and hybridized with a sequencing primer for region of interest. The first dNTP, for example dATP, is

added to the reaction and DNA polymerase catalyzes the insertion of dNTP at the sequence, if it is complementary to the nucleotide in the template strand. If the dNTP is incorporated, PPi will be released. Afterwards ATP sulfurylase converts PPi to ATP in the presence of APS. ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates a visible light signal equivalent to the amount of ATP. Afterwards, unincorporated nucleotides and ATP will be degraded by apyrase followed by adding another nucleotide and starting a new cycle of the reaction (figure 15).

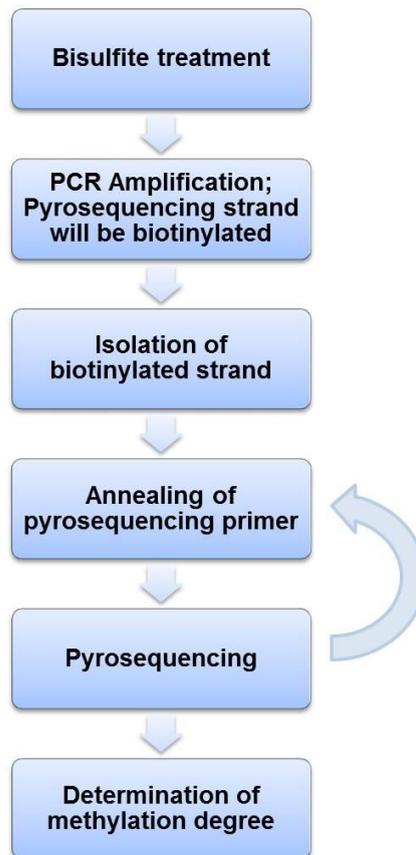


Figure 15: Overview of pyrosequencing workflow.

The light signal will be detected by CCD sensors and can be seen as a peak in the raw data output. The height of each peak is proportional to the number of nucleotides incorporated. A target region of up to 350 bp can be analyzed. For this project the pyrosequencing method provided by Qiagen was used.

2.6.2.2. Laboratory procedure

Primer for PCR amplification and pyrosequencing were synthesized by Biotex. 600 ng of genomic DNA were treated with sodium bisulfite using the EpiTect[®] 96 Bisulfite according to the manufacturer's instructions. Regions of interest were amplified using 25 ng of bisulfite treated human genomic DNA and 5 pmol of forward and reverse primer, one of them was biotinylated.

PCR primer mix included the following additional reagents in a total volume of 25 µl:

1x HotStar Taq buffer
 1.6 mM MgCl₂
 100 µM dNTPs
 4 U HotStar Taq polymerase

Afterwards the DNA was amplified using the following PCR protocol (table 4).

Table 4: Cycling protocol for PCR amplification for pyrosequencing process.

PCR amplification		
Temperature	Duration	Number of cycles
95°C	15 min	1
95°C	30 sec	50
56*	30 sec	
72°C	20 sec	
72°C	5 min	1
4°C	∞	

* may be adapted based on amplicon temperature

After verification by gel electrophoresis on 2% agarose gel, 5 µl of PCR product with 40 µl binding buffer were incubated for 10 min at RT with shaking.

Binding buffer included the following components:

10 mM Tris
 2 M NaCl
 1 mM EDTA
 0.1 % Tween 20; pH 7.6; adjusted with 1 M HCl

4 µl of streptavidin coated sepharose beads and 33 µL of ddH₂O were added. The binding mix was purified and rendered single-stranded using the Pyrosequencing

Vacuum Prep Workstation according to the manufacturer's instructions. Beads were released into 12 µl annealing buffer with the following components:

20 mM Tris

4 mM Mg-Acetate; pH 7.6; adjusted with 4 M acetic acid

4 pmol of the respective sequencing primer

Primers were annealed to the target by incubating at 80°C for 2 min. Quantitative DNA methylation analysis was carried out on a PSQ 96MD system with the PyroGold SQA Reagent Kit and results were analyzed using the Q-CpG software (V.1.0.9, Pyrosequencing).

2.7. Statistical analysis

2.7.1. Preprocessing and QC

2.7.1.1. Genome-wide DNA Methylation analysis

The two different chemistry technologies Infinium I and Infinium II used on the Infinium HumanMethylation450 BeadChip show a shift in the density curves between the probes detected by both chemistries [131, 134]. Therefore, it is recommended to normalize data prior to analysis so that the methylation values detected by the two chemistries are comparable. Different normalizations have been developed to this end. In the presented projects two methods, i.e. subset quantile normalization (SQN) [134] and beta mixture quantile normalization (BMIQ) [135] were applied to the data from the nested case-control study and the cross-sectional study, respectively.

After a first quality check raw methylation data were extracted with Illumina® GenomeStudio Version 2011.1 (respectively 2010.3; see chapter 2.5.3.), Methylation Module 1.9.0 (respectively 1.8.5; see chapter 2.5.3.) and preprocessed using SQN and BMIQ for normalization, and the packages *lumi* (version 2.8.0) and *methylumi* (version 2.2.0) using R (version 2.15.1 [136]). The pipelines cover the following preprocessing steps: First, probes with signals being summarized from less than three functional beads, and probes associated with a detection p-value larger than 0.01 were defined as low-confidence probes. Samples with more than

5% low-confidence probes were removed from the analysis. Second, probes representing or being located in 50 bp proximity to SNPs with minor allele frequency at least 5% were excluded from the data set to avoid confounding of the methylation level by genetic variation. Third, color bias adjustment using smooth quantile normalization, separately for the two color channels, and background correction based on the negative control probes present on the Infinium HumanMethylation450 BeadChip were conducted using the R package *lumi*. Finally, the pipeline provides a SQN and BMIQ normalization step, respectively. Following normalization, methylation values of low-confidence probes associated were set to missing, and probes with more than 5% low-confidence signals were removed from the analysis. Potential technical bias due to plate and chip assignment of the samples can be neglected for nested case-control study as the matched cases and controls were located next to another on the Infinium HumanMethylation450 BeadChip as well as on the same bisulfite conversion plate.

Due to missing values for covariables, four matched pairs had to be excluded for the nested case-control discovery study. Finally, 196 case and controls each were included in the statistical analysis (table 5). These data are also included in the replication stage of the LOLIPOP study.

Table 5: Number of probes and samples excluded during preprocessing.

	SQN (nested case-control study)		BMIQ (cross-sectional study)	
	Matched pairs (number)	Probes (number)	Samples (number)	Probes (number)
total	200 matched pairs	485,577	1,814	485,577
excluded	4	29,664	15	44,025
final	196	455,913	1,799	441,552

After finishing the analysis, it was found out that not all probes including a SNP are excluded based on the list provided by Touleimat and Tost [134] during the normalization method. Therefore, for the top hits, for potential SNPs using lists provided by Chen *et al.* [137] as well as Price *et al.* [138] were referred to and the minor allele frequency (MAF) including the European population [Utah Residents

(CEPH) with Northern and Western European ancestry (CEU) (n=99), Finnish in Finland (FIN) (n=99), Toscani in Italia (TSI) (n=107), British in England and Scotland (GBR) (n=91), Iberian population in Spain (IBS) (n=107)], via 1000 genomes (http://browser.1000genomes.org/Homo_sapiens/UserData/Allele?db=core; phase 3) was calculated. All SNPs included in both lists for the found CpG sites have a $MAF \leq 0.05$, therefore the CpG sites were not excluded.

2.7.1.2. Replication and Fine Mapping

Plate respectively SpectroChip effects were not investigated as matched cases and controls were localized next to another on each plate, and can therefore be neglected. Five matched pairs were excluded as the cases' time to diabetes diagnosis was less than 1 year after the baseline time point, and therefore, it cannot be excluded that they were already diabetic at this time point. A further matched pair has to be excluded due to missing covariable values. Finally, 790 samples (395 matched cases and controls) were included in the replication study (figure 8).

2.7.1.3. Comparison SQN vs BMIQ

As mentioned previously, normalization of raw methylation data is strongly needed to make the signals detected via Infinium I and Infinium II comparable. In the following section, the two normalization methods used for this doctoral thesis, SQN and BMIQ, are described in more detail. Furthermore, it will be pointed out why a new normalization method was used for the second part of this project.

SQN normalized Infinium II signals using Infinium I as anchors, conducted separately for six probe categories, defined by their relative position to the CpG islands. Therefore, the Infinium I/Infinium II bias was corrected, while keeping the relative distribution across CpG islands as well as shores, shelves, and distant related probes of Infinium II signal. SQN computes reference quantiles from Infinium I probes and uses them to normalize Infinium II probes [134].

The BMIQ pipeline uses the statistical distribution characteristics of type 1 probes and uses them to adjust the β -values of type 2 probes. 3-state beta-mixture models were used to assign probes to methylation states followed by transformation of probabilities into quantiles and last – to preserve the monotonicity and continuity of data – a methylation dependent dilation transformation [135].

At the time when the nested case-control study was processed, SQN appeared to be the best publically available strategy to normalize methylation data. During subsequent months, studies were published comparing the normalization methods. Marabita and colleagues compared different published normalization pipelines and found that BMIQ reduced the Infinium I/Infinium II probe type biases effectively. In addition, technical variability of almost all investigated pairs was reduced. Furthermore, they show that while SQN reduced the bias between probe types, technical variability was amplified when this method was used [139]. Due to the advanced process of the project and started replication it was decided to use the SQN pipeline for the studies of association of DNA methylation and incident T2D and switch to the BMIQ pipeline for the other project.

2.7.2. Statistical analysis

2.7.2.1. Nested case-control study

392 matched samples (196 cases and controls each) and 790 matched samples (395 cases and controls each) were included in the discovery and replication studies, respectively (figure 8).

Associations between DNA methylation at baseline (methylation β -values) and T2D development during follow-up were determined using conditional logistic regression models with the R package *survival*, version 2.37.4 [140]. Conditional logistic regression was used to account for the matched design of nested case-control studies, thereby allowing for unbiased effect estimation. Different models were fitted (table 6).

Table 6: Regression models for methylome-wide association study by incident T2D.

Model	Adjustment variables
M1	no adjustment (note that age and sex is accounted for by 1:1 matching already)
M2	BMI

CpG sites showing genome-wide association with incident T2D in model 1 ($p < 0.05$ after Benjamini-Hochberg (B-H) correction for multiple testing) were selected for replication. cg22800477 (annotated to *CASZ1*) and cg09154213 (annotated to *TMEM57*) could not be included in analysis via EpiTYPER[®] as no primer could be designed due to technical issues. Additionally, two candidate CpG sites annotated to *TCF7L2* and *CDKAL1* [37] were embedded, that were among the 1,000 best hits for different adjustment models. CpG sites annotated to *TMEM57* and *CASZ1* were replicated using pyrosequencing. Data were analyzed in analogy to the discovery study. In the replication study, p-values were corrected for multiple testing using the conservative Bonferroni method, accounting for eight independent CpG sites.

2.7.2.2. Replication for LOLIPOP study

Data from the nested case-control study were inquired for replication of methylome-wide significant hits from the LOLIPOP study which investigated the prediction of incident T2D through DNA methylation markers in peripheral blood amongst Indian Asians and Europeans. The discovery study comprises 1,074 incident T2D patients and 1,590 controls with Indian Asian ancestry. The replication panel included 377 incident cases and 764 matched controls with European ancestry and 647 incident cases and 1,073 controls with Indian Asian ancestry. In the following section only the statistical part for the KORA study is described. For this study BMIQ-normalized methylation β -values of 196 matched cases and controls from the nested case-control study were used. The different adjustment models are presented in table 7. For the calculation of a methylation score, the discovered methylation markers were standardized (z-transformed) and weighted by their discovery effect sizes.

Statistical analysis was done as described under 2.7.2.1. In addition, mean and standard deviation of methylation markers and methylation score were determined for T2D cases and controls.

Table 7: Regression models for replication of the methylome-wide association study by incident T2D (LOLIPOP study).

Model	Adjustment variables
M1	no adjustment
M2	age
M3	M2 + waist hip ratio + BMI

2.7.2.3. Cross-sectional study

1,448 non-diabetic participants from the follow-up study F4 (see chapter 2.2.2; figure 9) were included in the methylome-wide association study for fasting glucose, 2-hour glucose, fasting insulin, and HbA1c as well as 617 for 2-hour insulin, respectively and 1,440 for HOMA-IR. Quantitative traits, excluding HbA1c and 2-hour insulin, were log transformed to obtain an approximate normal distribution. The association between DNA methylation and T2D-related quantitative traits (fasting glucose, 2-hour glucose, fasting insulin, HOMA-IR, and HbA1c) were assessed using linear mixed effect models (R package *nlme* [141]) with trait as response variable and methylation β -value as independent variable, including additional covariates as described in table 8 and accounting for plate effects.

Table 8: Regression models for methylome-wide association study by measures of glucose metabolism.

Model	Adjustment variables
M1	Age, sex + estimated white blood cell proportions [142]
M2	M1 + BMI

Furthermore, 2-hour insulin (cube root transformed) was analyzed similarly including the CpG sites which showed genome-wide significance in model 1 with phenotypes named above due to the small number of samples with available 2-hour insulin data. Covariates described in table 8 were included for statistical analysis. All results were corrected for multiple testing using the Benjamini-Hochberg (B-H) method. Results were defined as significant where B-H-adjusted $p < 0.05$. Results are presented with

coefficient (expressing how many SD a dependent variable will change, per SD increase in the predictor variable), p-value and B-H adjusted p-value.

Due to the fact that not all probes including a SNP, which could be problematic as they can influence the binding of probes on the beads labeled at the array, were excluded via a list provided by Touleimat and Tost [134] during the normalization method, significant CpG sites were checked for potential existence of SNPs. The lists provided by Chen *et al.* [137] as well as Price *et al.* [138] were used and the MAF based on the data from the five European populations [CEU (n=99), FIN (n=99), TSI (n=107), GBR (n=91), IBS (n=107)] included in the 1000 genome project (http://browser.1000genomes.org/Homo_sapiens/UserData/Allele?db=core;phase=3) was calculated. All SNPs included in the lists of the detected CpG sites have a MAF ≤ 0.05 , therefore they were not excluded.

For sensitivity analyses, samples were stratified according to DNA methylation quintiles at the significant CpG sites. Mean \pm SD was calculated. P-values were determined using a linear regression, determining the regression between the phenotype and the median of the methylation-quintile, for continuous and χ^2 test for categorical variables.

Statistical analysis were performed using R (version 2.15.3 or higher).

2.7.2.4. Gene expression analysis

Total RNA was extracted from whole blood under fasting conditions according to the manufacturer's instructions using the PAXgene Blood miRNA Kit (Qiagen). Gene expression profiling was carried out as described elsewhere using the Illumina Human HT-12 v3 Expression BeadChip [143].

The samples for gene expression analysis were preprocessed as described elsewhere [143]. For a subset of 533 participants, both methylation and gene expression data were available (see also chapter 2.2.2.). In this subset the associations between gene expression and DNA methylation and the associations between gene expression and phenotype were analyzed using linear regression

models. Transcript probes that mapped in a ± 500 kb window around the CpG site were included. Analyses were adjusted for technical variables [sample storage time, RNA integrity number (RIN), RNA amplification batch] [143], age, sex, BMI, and estimated white blood cell proportions [142] and plates. Multiple testing was accounted for using the B-H procedure. Results were defined as significant where B-H-adjusted $p \leq 0.05$.

All statistical analyses were performed using R (version 2.15.3 or higher).

2.8. Ingenuity Pathway analysis

The Ingenuity Pathway Analysis (IPA) software was used to detect potential pathways and networks in the DNA methylation data relevant in the nested case-control study and cross-sectional study in an unbiased way. The 1,000 CpG sites with the smallest p-values assessed in the different analyses were included.

In the case that two or more genes were annotated to one CpG site, the CpG was set for each annotated gene alone. CpG name, gene name, p-value, and beta were uploaded to IPA and each identifier was mapped to the corresponding term (ID: gene name; others: beta; p-value: p.value). Ingenuity Knowledge Base (Gene only) and Ingenuity Knowledge Base (Gene + Endogenous Chemicals) were used for nested case-control study and cross-sectional study, respectively. Canonical pathways were investigated and pathways with a p-value < 0.05 after B-H correction were defined as statistically significant.

2.9. Material

2.9.1. Laboratory equipment

384-well plate	Thermo Fisher Scientific (Waltham, MA, USA)
96-well plate	Thermo Fisher Scientific (Waltham, MA, USA)
Aluminium seal	Thermo Fisher Scientific (Waltham, MA, USA)
Cap Mat	Thermo Fisher Scientific (Waltham, MA, USA)
Centrifuge	Sigma 4K15C (Sigma Laborzentrifugen, Osterode, Germany) Rotanta 46 RS (Hettich, Tuttlingen, Germany) Mikrozentrifuge (NeoLab, Heidelberg, Germany) Centrifuge 5417R (Eppendorf AG, Hamburg, Germany)
Clear Spacer	Tecan AG (Crailsheim, Germany)
Falcons 15 ml/ 50 ml	BD (Heidelberg, Germany)
Falcons 15 ml/ 50 ml	Sarstedt (Nuernbrecht, Germany)
Felix 2000	Biostep (Wolferstadt, Germany)
Flow-Through Chambers	Tecan AG (Crailsheim, Germany)
Gel electrophoreses chamber	Biozym (Oldendarf, Germany)
Gel electrophoresis device: Bio-Rad Power Pac 300/3000	BIO-RAD Laboratories (Munich, Germany)
Glass back plate	Illumina (San Diego, CA, USA)
Heat block	SciGene (Sunnyval, CA, USA)
Heat oven	Memmert (Schwabach, Germany)
Heat Sealer	Thermo Fisher Scientific (Waltham, MA, USA)
Hyb Chamber including BeadChip Hyb Chamber gasket Hyb Chamber insert	Illumina (San Diego, CA, USA)
Hybridization oven	Illumina (San Diego, CA, USA)
iScan	Illumina (San Diego, CA, USA)
Kimwipes	Kimberly-Clark (Koblenz-Reinh., Germany)
MALDI-TOF MS	Agena Bioscience formerly Sequenom (Hamburg, Germany)
Mass Array™ Nanodispenser	Agena Bioscience formerly Sequenom (Hamburg, Germany)
Multi-sample BeadChip alignment fixture including Black frames	Illumina (San Diego, CA, USA)

Metal clamps	
NanoDrop (8-Sample Spectrophotometer)	PeqLab (Erlangen, Germany)
PCR cycler	Thermal Cycler C1000™ (BIO-RAD Laboratories, Munich, Germany) DNA Engine Tetrad (MJ Research, South San Francisco, USA)
Single pipettors (Rainin) 1–10 µl, 5–50 µl, 50–300 µl, 100 – 1,000 µl	Mettler Toledo (Gießen, Germany)
Multi-channel pipettes (Rainin) 1–10 µl, 5–50 µl, 50–300 µl, 100 – 1200 µl	Mettler Toledo (Gießen, Germany)
Pipettes tips	Mettler Toledo (Gießen, Germany)
PSQ 96MD system	Qiagen (Hilden, Germany)
PyroMark Q96 HS Nucleotide Tip	Qiagen (Hilden, Germany)
PyroMark Q96 HS Q96 Reagent Tip	Qiagen (Hilden, Germany)
Pyrosequencing Vacuum Prep Workstation	Pyrosequencing AB (Uppsala, Sweden)
Robot BeadChip Alignments Fixtures	Illumina (San Diego, CA, USA)
Robot Tecan Genesis RSP 150	Tecan AG (Crailsheim, Germany)
Robot Tip Alignment Guide	Illumina (San Diego, CA, USA)
Thermo Seal slide	Thermo Fisher Scientific (Waltham, MA, USA)
Seal slide	Qiagen (Hilden, Germany)
Taq polymerase	Qiagen (Hilden, Germany)
Tube racks for vacuum desiccator	VWR (Darmstadt, Germany)
Vacuum desiccator + exhaustor (Nalgene)	Thermo Fisher Scientific (Waltham, MA, USA) KNF Laborport (Freiburg, Germany)
High Speed microplate shaker	Illumina (San Diego, CA, USA)
Vortex mixer	Velp Scientifica (Bohemia, NY, USA)
Rotilabo® Liquid reservoir PVC	Carl Roth (Karlsruhe, Germany)
Water circulator	VWR (Darmstadt, Germany)

2.9.2. Chemicals, reagents, enzymes and assays

0.1N NaOH (Sodium hydroxide)	Sigma Aldrich (Taufkirchen, Germany)
0.5M EDTA	Merck (Darmstadt, Germany)
100% Ethanol	Merck (Darmstadt, Germany)
100% Isopropanol	Merck (Darmstadt, Germany)
4-point calibrant for EpiTYPER[®]	Agena Bioscience formerly Sequenom (Hamburg, Germany)
6x Loading buffer	VWR (Darmstadt, Germany)
95% formamide/1mM EDTA	Carl Roth (Karlsruhe, Germany)
Agarose powder	Biozym (Oldendorf, Germany)
Amicroparticle enzyme immunoassay	Abbott Laboratories (Ludwigshafen, Germany)
Boric Acid	Carl Roth (Karlsruhe, Germany)
Buffer and MgCl₂ for PCR	Qiagen (Hilden, Germany)
Deionized water	Merck (Darmstadt, Germany)
Dextro OGT	Boehringer (Mannheim, Germany)
DNA Ladder	Thermo Fisher Scientific (Waltham, MA, USA)
Lamda DNA/EcoRI + HindIII Marker	
GeneRuler 1kb DNA Ladder	
dNTP	Thermo Fisher Scientific (Waltham, MA, USA)
EDTA	Merck (Darmstadt, Germany)
Glucosquant	Roche (Diagnostics, Mannheim, Germany)
HotStarTaq DNA Polymerase	Qiagen (Hilden, Germany)
Human Methylation Controls - Mix	EpigenDX (Hopkinton, MA, USA)
Midori Green	Nippon Genetics (Dueren, Germany)
Nanopure H₂O	Carl Roth (Karlsruhe, Germany)
Primer for PCR and EpiTYPER[®]	Metabion (Planegg, Germany)
Primer for pyrosequencing	Biotez (Berlin, Germany)
PyroMark Gold Q96 SQA Reagents	Qiagen (Hilden, Germany)
Streptavidin coated sepharose beads	GE Healthcare (Uppsala, Sweden)
Taq-Polymerase + buffer for PCR	Qiagen (Hilden, Germany)
Tris	AppliChem, Inc. (St. Louis, MI, USA)

2.9.3. Kits

EpiTect® 96 Bisulfite Kit including Bisulfite Solution DNA Protect Buffer Carrier RNA Buffers EpiTect 96-well plates	Qiagen (Hilden, Germany)
EpiTYPER® T Reagent Set (10x384) including Complete PCR Reagent Set MassCLEAVE T7 Kit (T Cleavage) Clean Resin Kit SpectroCHIP II Arrays	Agena Bioscience formerly Sequenom (Hamburg, Germany)
EZ-96 DNA Methylation™ Kit (Shallow-Well Format) including Bisulfite conversion plate Collection plate Binding plate Elution plate Bisulfite conversion reagent Wash buffer Dilution buffer Binding buffer Desulphonation buffer Elution buffer	Zymo Research (Irvine, CA, USA)
Infinium HumanMethylation450 BeadChip including MSA1(multi sample amplification 1 mix) RPM (random primer mix) MSM (multiple sample amplification master mix) ..FMS (fragmentation solution) PM1 (precipitation solution) RA1 (resuspension, hybridization, and wash solution) PB2 (humidifying buffer used during hybridization) XC1 (XStain BeadChip solution 1) XC2 (XStain BeadChip solution 2) XC3 (XStain BeadChip solution 3) XC4 (XStain BeadChip solution 4) STM (superior two-color master mix)	Illumina, Inc. (San Diego, CA, USA)

ATM (anti-stain two-color master mix)
TEM (two-color extension master mix)
PB1 (reagent used to prepare BeadChip for hybridization)
HumanMethylation450 BeadChip

2.9.4. Computer software and programs

Illumina GenomeStudio (Methylation Module)	Illumina Inc. (San Diego, CA, USA)
MassARRAY EpiTyper 1.2 EpiDesigner.com Plate Editor Analyzer	Agena Bioscience formerly Sequenom, Hamburg, Germany
MassArray package	Thompson and Grealley (2009). MassArray: Analytical Tools for MassArray Data. R package version 1.10.0.
Q-CpG software (V.1.0.9, Pyrosequencing)	Qiagen (Hilden, Germany)
R software 2.15.3	www.r-project.org
RT Workstation 3.4: Chip Linker Caller Acquire	Agena Bioscience formerly Sequenom, Hamburg, Germany

2.9.5. Online databases and programs

1000 Genome	http://browser.1000genomes.org/
Ingenuity pathway analyses (Qiagen)	http://www.ingenuity.com/products/ipa
National Center of Biotechnology Information	http://www.ncbi.nlm.nih.gov/
R	http://www.r-project.org/
Sequenom`s Primer Design	www.epidesigner.com
UCSC Genome Browser	https://genome.ucsc.edu/index.html

3. Results

3.1. Association with incident T2D

For the analysis of methylome-wide association with incident T2D a nested case-control study was designed using samples from the baseline studies S3 and S4 in a discovery and K12 and S2 in a replication panel.

3.1.1. Study characteristics

Study characteristics for the discovery and the replication study are shown in table 9 and table 10, respectively. 54.1% of cases and controls were male and the mean ages of all cases and controls included were 57.8 and 57.6, respectively. There were no statistically significant differences between the two groups with respect to age and sex. However the BMI was significantly increased in cases compared to controls. At the same time white blood cell count, C-reactive protein (CRP), HbA1c, and systolic blood pressure were also significantly increased. HDL cholesterol was significantly decreased in cases compared to controls. Furthermore, cases were significantly more inactive and had a higher tendency to get a myocardial infarct. In addition the positive family history and smoking status was significantly different in people with incident T2D (table 9).

The mean age of the samples included in the replication study was 56.8 years for cases and 56.7 for controls and thus comparable to the discovery study. BMI, systolic blood pressure and total cholesterol were significantly increased in cases, whereas HDL cholesterol was significantly decreased. The study population included slightly more males than females. Cases were significantly more inactive, had a higher T2D family history, and a distinct smoking behavior. CRP, white blood cell count, and HbA1c were not available in the replication study (table 10).

Table 9: Study characteristics of the discovery study – a total of 196 matched case-controls pairs.

	Cases	Controls	
	Mean (SD)/%	Mean (SD)/%	p-value
Sex [% male] #	54.1	54.1	1
Age [years] *	57.8 (8.9)	57.6 (8.8)	0.20
BMI [kg/m²] *	30.9 (4.8)	27.5 (4.0)	1.1 x 10 ⁻¹⁴
HbA1c [%] *	5.8 (0.8)	5.3 (0.4)	3.1 x 10 ⁻¹⁴
C-reactive protein [mg/l] *	3.8 (5.5)	2.2 (3.7)	1.4 x 10 ⁻⁴
White blood cell count [nl] *	7.5 (2.7)	6.6 (1.7)	2.0 x 10 ⁻⁵
Total cholesterol [mg/dl] *	237.9 (41.9)	242.9 (42.9)	0.25
HDL cholesterol [mg/dl] *	47.0 (13.2)	55.4 (16.9)	1.2 x 10 ⁻⁸
Systolic blood pressure [mm Hg] *	140.3 (19.3)	135.0 (20.2)	7.4 x 10 ⁻³
Alcohol consumption [g/d] *	17.3 (25.4)	19.2 (26.3)	0.40
Smoking status [%] #			
current	25.0	14.3	1.3 x 10 ⁻²
ex	34.7	35.2	
never smoker	40.3	50.5	
Physically active [%] # (combination of activity during summer and winter)	37.8	50.0	1.1 x 10 ⁻²
Myocardial infarction [%] #	4.6	1.5	5.0 x 10 ⁻²
Parental T2D history [%] #			
yes	34.2	20.4	7.6 x 10 ⁻³
no	41.3	53.6	
don't know	24.5	26.0	

p-values were determined through likelihood-ratio tests.

* Data are presented as medians for the continuous variables.

Data are presented as the number in percentage for categorical variables

Table 10: Study characteristics of the replication study - a total of 395 matched case-controls pairs.

	Cases	Controls	p-value
	Mean (SD)/%	Mean (SD)/%	
Sex [% male] #	60.3	60.3	1
Age [years] *	56.8 (9.7)	56.7 (9.8)	0.42
BMI [kg/m²] *	29.7 (4.2)	27.1 (3.5)	5.7x10 ⁻²⁰
C-reactive protein [mg/l] *	3.8 (6.5)	2.9 (4.6)	0.19
Total cholesterol [mg/dl] *	252.8 (45.4)	246.6 (42.3)	4.1 x 10 ⁻²
HDL cholesterol [mg/dl] *	50.5 (15.0)	57.9 (16.7)	2.4 x 10 ⁻¹¹
Systolic blood pressure [mm Hg] *	140.8 (18.9)	134.3 (18.6)	1.0 x 10 ⁻⁶
Alcohol consumption [g/d] *	21.0 (27.2)	19.1 (24.4)	0.22
Smoking state [%] #			
current	25.3	17.7	
ex	33.7	35.7	2.4 x 10 ⁻²
never smoker	41.0	46.6	
Physically active [%] † (combination of activity during summer and winter)	27.9	34.7	3.4 x 10 ⁻²
Myocardial infarct [%] #	2.5	2.8	0.81
Parental T2D history [%] #			
yes	26.3	20.3	
no	47.3	60.0	1.9 x 10 ⁻³
don't know	25.8	19.2	

p-values were determined through likelihood-ratio tests.

* Data are presented as medians for the continuous variables.

Data are presented as the number in percentage for categorical variables

3.1.2. Genome-wide DNA methylation analysis

Results for the methylome-wide association analysis with incident T2D are displayed in table 11. In the first adjustment model (without adjustment) the methylation degrees of six CpG sites [cg11057824 (*C140rf182*), cg18514820 (*VIM*), cg20587409 (*CDKAL1*), cg22876894 (*not annotated*), cg23951816 (*TCF7L2*), cg25333225 (*AKT2*)] reached genome-wide significance after correction for multiple testing (B-H adjusted p-values 2.1 x 10⁻² - 6.8 x 10⁻³). After additional adjustment for BMI (adjustment model 2) as a high risk factor for T2D, no genome-wide significant CpG sites could be detected. Results of the methylome-wide analysis are shown as

Manhattan plots for models 1 and 2 (figure 16). Annotation information of detected CpG sites is summarized in appendix table 3. Furthermore, a slight, but significant, difference for the methylation degree for the aforementioned CpG sites between cases and controls was found (p-values 10^{-7} - 10^{-8}). Except cg11057824, the methylation degrees of the cases were increased at the CpG sites compared to controls (figure 17).

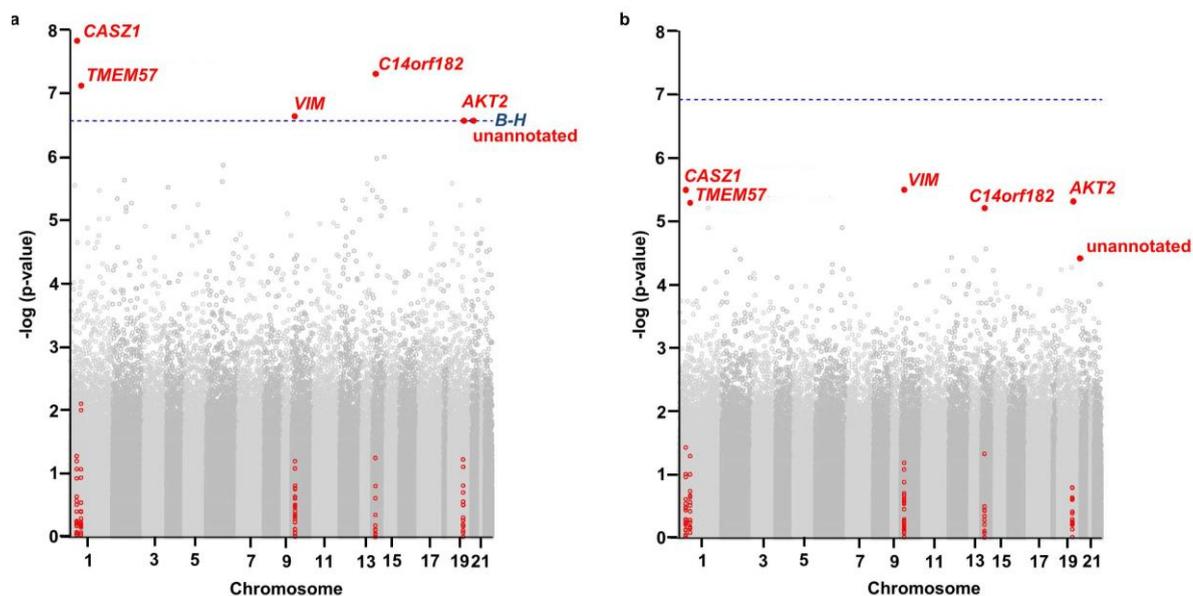


Figure 16: Results of methylome-wide association analysis with incident T2D. a) Results without adjustment. b) Results after adjustment for BMI; B-H: Benjamini-Hochberg

Table 11: Results of methylome-wide association analysis with incident T2D for the discovery study.

CpG	Gene	Chrom.	Without adjustment			Adjusted for BMI		
			OR (95% CI)	p-value	B-H adjusted p-value	OR (95% CI)	p-value	B-H adjusted p-value
cg09154213	<i>TMEM57</i>	1	3.18 (2.09, 4.85)	7.8×10^{-8}	1.1×10^{-2}	2.66 (1.75, 4.06)	5.2×10^{-6}	0.43
cg22800477	<i>CASZ1</i>	1	1.49 (1.30, 1.71)	1.5×10^{-8}	6.2×10^{-3}	1.45 (1.24, 1.70)	3.2×10^{-6}	0.43
cg18514820	<i>VIM</i>	10	2.54 (1.78, 3.61)	2.4×10^{-7}	1.9×10^{-2}	2.58 (1.73, 3.84)	3.3×10^{-6}	0.43
cg11057824	<i>C14orf182</i>	14	0.92 (0.89, 0.95)	4.9×10^{-8}	1.0×10^{-2}	0.92 (0.89, 0.96)	6.2×10^{-6}	0.43
cg25333225	<i>AKT2</i>	19	2.86 (1.91, 4.26)	2.7×10^{-7}	1.9×10^{-2}	2.90 (1.84, 4.59)	4.9×10^{-6}	0.43
cg22876894	unannotated	20	2.57 (1.80, 3.69)	2.7×10^{-7}	1.9×10^{-2}	2.29 (1.54, 3.41)	3.9×10^{-5}	0.93

Chrom.=chromosome; OR: odds ratio; CI: confidence interval; B-H: Benjamini-Hochberg

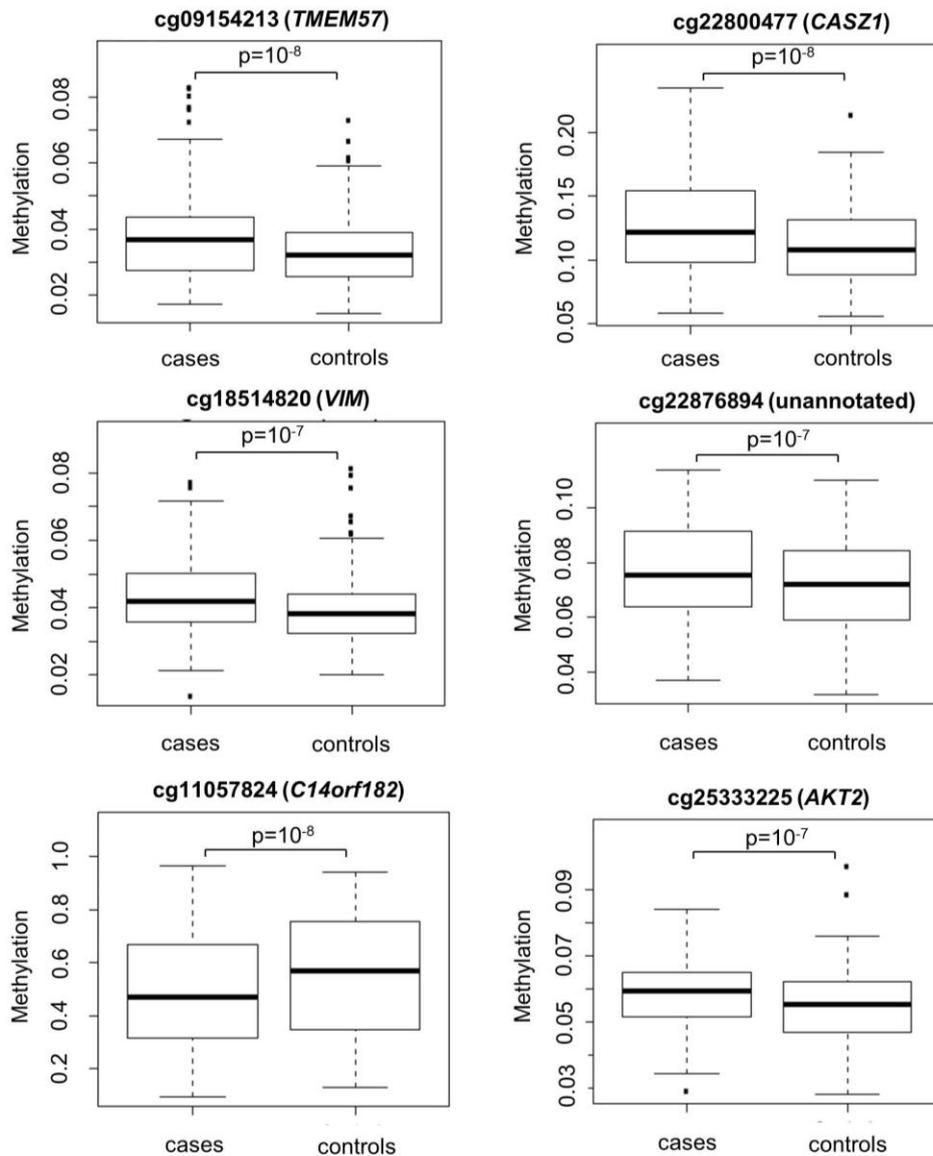


Figure 17: Differences for methylation degree of methylome-wide CpG sites between cases and controls.

3.1.3. Replication and fine mapping

To assess and validate the reliability of the CpG sites identified as associated with incident T2D in the discovery study, hits were replicated in an independent study using EpiTYPER[®] and pyrosequencing. Results for the replication study are shown in table 12. As it was not possible to analyze *CASZ1* and *TMEM57* using EpiTYPER[®] due to technical problems they were replicated using pyrosequencing. Additionally, two candidate CpG sites annotated to *TCF7L2* and *CDKAL1* [37] were embedded in analysis using EpiTYPER[®]. CpG sites cg20587409 (*CDKAL1*) and

cg23951816 (*TCF7L2*) were included as *TCF7L2* is one of the genes most strongly associated with T2D to date and *CDKAL1* is a well-established locus as well. Both CpG sites can also be found within the top 1,000 CpG sites associated with incident T2D in the discovery study.

The methylation degree of none of the leading CpG sites was significantly associated with incident T2D neither with nor without adjustment for BMI after correction for multiple testing using the Bonferroni method. However, the methylation at four CpG sites flanking the leading CpG site from the discovery study showed significant association with incident T2D in the replication study. Three of them belong to the *AKT2* amplicon (CpG site 1, 2/3, 16) and one to *C14orf182* (CpG site 10). CpG site 10 of the *C14orf182* amplicon and CpG site 2/3 were still significant after adjustment for BMI (table 13). The fragments including CpG sites 1 and 16 of the *AKT2* amplicon contains the same mass. Therefore, it was not possible to distinguish between these two results, hence interpretation of results is limited.

CpG site 10 of *C14orf182* is located 151 bp downstream from the leading CpG site. CpG sites 1 and 2/3 of the *AKT2* amplicon are 177 bp and 169/162 bp downstream from the leading CpG site, whereas CpG site 16 is 43 bp upstream (figure 18). Furthermore, CpG sites 1 and 16 of the *AKT2* amplicon are also included on the Infinium HumanMethylation450 BeadChip, namely cg14260485 and cg03023952, having uncorrected p-values of 0.79 and 0.54 respectively in the discovery study without adjustment. The CpG sites have correlations between 0.089 and 0.231 with the leading CpG site.

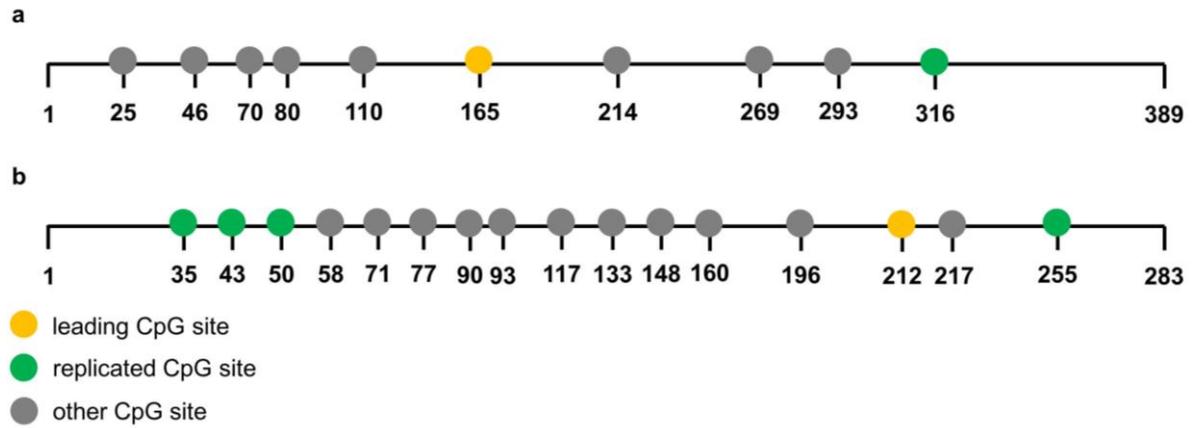


Figure 18: Positions of CpG sites a) cg11057824 (annotated to *C14orf182*) and b) cg25333225 (annotated to *AKT2*) within the amplicon, respectively. The number presents the number of bases within the amplicon.

Table 12: Results of methylome-wide analysis with T2D for the replication study showing the leading CpG sites from the discovery study.

Amplicon	Lead CpG	Gene	CpG in Amplicon	Chrom.	Without adjustment			Adjusted for BMI		
					OR (95% CI)	p-value	Bonferroni adjusted p-value	OR (95% CI)	p-value	Bonferroni adjusted p-value
TMEM57	cg09154213	<i>TMEM57</i>	CpG_5	1	1.00 (0.90, 1.12)	0.95	1	1.00 (0.88, 1.13)	0.95	1
CASZ1	cg22800477	<i>CASZ1</i>	CpG_3	1	1.01 (0.94, 1.08)	0.83	1	1.00 (0.93, 1.08)	0.97	1
A5a_cg20587409_5	cg20587409	<i>CDKAL1</i>	CpG_1	6	1.02 (0.97, 1.07)	0.52	1	1.01 (0.95, 1.07)	0.74	1
A2a_cg18514820_1	cg18514820	<i>VIM</i>	CpG_17.18	10	0.97 (0.89, 1.04)	0.37	1	0.95 (0.88, 1.04)	0.30	1
A6a_cg23951816_9	cg23951816	<i>TCF7L2</i>	CpG_1.2	10	1.01 (0.95, 1.07)	0.69	1	1.05 (0.99, 1.13)	0.13	1
A1a_cg11057824_13	cg11057824	<i>C14orf182</i>	CpG_6	14	0.98 (0.96, 1.00)	9.8×10^{-3}	7.9×10^{-2}	0.98 (0.96, 1.00)	2.6×10^{-2}	0.21
A4b_cg25333225_4	cg25333225	<i>AKT2</i>	CpG_14	19	1.03 (0.89, 1.20)	0.70	1	1.04 (0.88, 1.23)	0.64	1
A3b_cg22876894_5	cg22876894	unannotated	CpG_7	20	0.99 (0.92, 1.06)	0.69	1	0.98 (0.91, 1.06)	0.59	1

Chrom.: chromosome; OR: odds ratio; CI: confidence interval

Table 13: Results of methylome-wide analysis with incident T2D for the replication study showing the significant CpG sites.

Amplicon	Lead CpG	Gene	CpG in Amplicon	Chrom.	Without adjustment			Adjusted for BMI		
					OR (95% CI)	p-value	Bonferroni adjusted p-value	OR (95% CI)	p-value	Bonferroni adjusted p-value
A1a_cg11057824_13	cg11057824	<i>C14orf182</i>	CpG_10	14	1.26 (1.08, 1.48)	4.4×10^{-3}	3.5×10^{-2}	1.28 (1.07, 1.53)	7.4×10^{-3}	5.9×10^{-2}
A4b_cg25333225_4	cg25333225	<i>AKT2</i>	CpG_1	19	0.70 (0.57, 0.87)	1.1×10^{-3}	8.7×10^{-3}	0.76 (0.60, 0.96)	2.4×10^{-2}	0.19
A4b_cg25333225_4	cg25333225	<i>AKT2</i>	CpG_2.3	19	0.60 (0.44, 0.81)	1.0×10^{-3}	8.4×10^{-3}	0.61 (0.43, 0.86)	5.4×10^{-3}	4.3×10^{-2}
A4b_cg25333225_4	cg25333225	<i>AKT2</i>	CpG_16	19	0.70 (0.57, 0.87)	1.1×10^{-3}	8.7×10^{-3}	0.76 (0.60, 0.96)	2.4×10^{-2}	0.19

Chrom.: chromosome; OR: odds ratio; CI: confidence interval

3.1.4. Pathway analysis

To functionally integrate the aforementioned results, pathway analyses were conducted using Ingenuity Software. Results for the first ten canonical pathways are presented in table 14 and table 15 for the discovery study without and with adjustment for BMI, respectively.

In total, two canonical pathways were found for incident T2D reaching the B-H corrected level of significance (p -values < 0.05) including the first ten pathways for the different adjustment models. For the first adjustment model no pathways were significantly enriched. In contrast, “G Beta Gamma Signaling” and “NGF Signaling” are the two top pathways for the adjustment model including BMI with B-H adjusted p -values 3.41×10^{-2} and 3.77×10^{-2} . In these pathways, 14 from a total of 117 and 15 from a total of 118 genes which were incorporated were amongst the top 1,000 CpG sites. “Integrin Signaling” and “Macropinocytosis Signaling” were detected in both adjustment models.

Table 14: Results of the pathway analyses for the discovery study without adjustment presenting the first ten canonical pathways.

	p-value	B-H adjusted p-value	Ratio
Virus Entry via Endocytic Pathways	3.23×10^{-4}	ns	13/99
Integrin Signaling	5.13×10^{-4}	ns	21/207
FAK Signaling	8.27×10^{-4}	ns	12/101
VEGF Family Ligand-Receptor Interactions	9.89×10^{-4}	ns	11/84
Neuregulin Signaling	1.02×10^{-4}	ns	12/102
Erythropoietin Signaling	1.31×10^{-4}	ns	10/78
Macropinocytosis Signaling	1.47×10^{-4}	ns	10/76
Glioma Signaling	1.99×10^{-4}	ns	12/112
Role of BRCA1 in DNA Damage Response	2.17×10^{-4}	ns	9/65
UCP-D-xylose and UDP-D-glucuronate Biosynthesis	2.32×10^{-4}	ns	2/7

B-H: Benjamini-Hochberg; ns: not significant after correction.

Table 15: Results for pathway analysis for the discovery study after adjustment for BMI presenting the first ten canonical pathways.

	p-value	B-H adjusted p-value	Ratio
G Beta Gamma Signaling	8.67×10^{-5}	3.41×10^{-2}	14/117
NGF Signaling	1.91×10^{-4}	3.77×10^{-2}	15/118
Integrin Signaling	6.09×10^{-4}	ns	21/207
Relaxin Signaling	8.32×10^{-4}	ns	16/159
Synaptic Long Term Depression	1.23×10^{-3}	ns	16/159
Macropinocytosis Signaling	1.62×10^{-3}	ns	10/76
Cardiac- β -adrenergic Signaling	1.83×10^{-3}	ns	15/154
AMPK Signaling	1.97×10^{-3}	ns	15/167
Phospholipase C Signaling	2.15×10^{-3}	ns	22/260
Gaq Signaling	2.22×10^{-3}	ns	16/168

B-H: Benjamini-Hochberg; ns: not significant

3.1.5. Replication for LOLIPOP study

Samples from the nested case-control study were included for replication in the investigation of the LOLIPOP study analyzing DNA methylation markers in peripheral blood and their prediction possibility of T2D amongst Indian Asians and Europeans. In the following chapter only the results are presented for which samples from the KORA study were included (see also chapter 10.1.).

In the discovery study for this project the methylation degrees of seven CpG sites were genome-wide significantly associated with incident T2D in Indian Asians with p-values $< 10^{-6}$ [cg19693031 (*TXNIP*), cg09152259 (*PROC*), cg04999691 (*C7orf29*), cg11024682 (*SREBF1*), cg18181703 (*SOCS3*), cg02650017 (*PHOSPHO1*), and cg06500161 (*ABCG1*)]. Five of these seven CpG sites were replicated combining data from Indian Asians as well as Europeans, namely cg19693031, cg11024682, cg18181703, cg02650017, cg06500161 with p-values < 0.05 (figure 19, table 16).

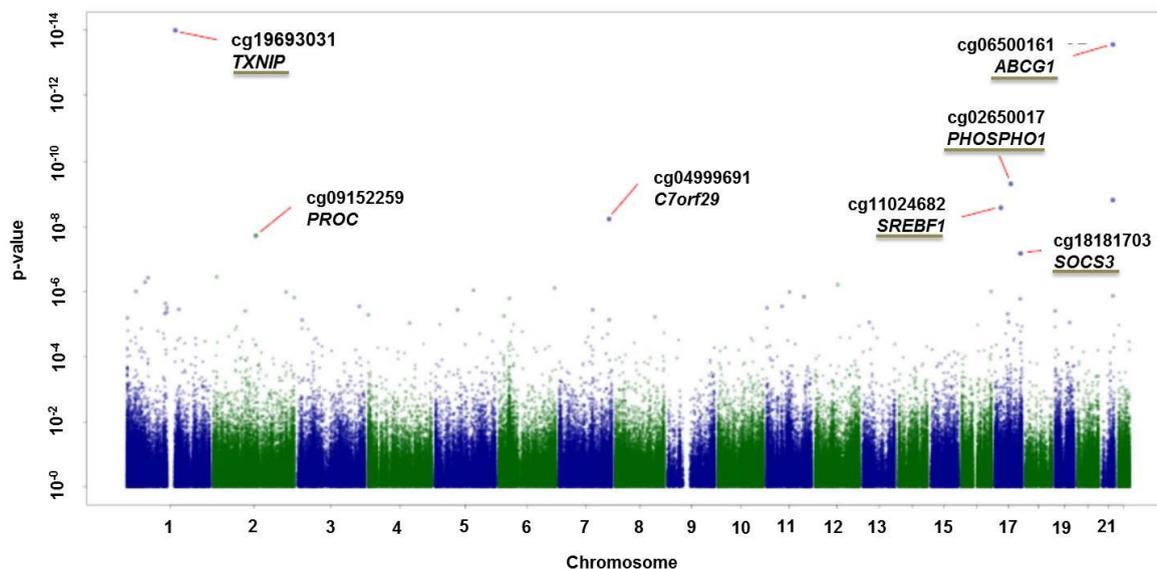


Figure 19: Results of methylome-wide association analysis with incident T2D. Underlined CpG sites were replicated with Europeans [modified after Chambers et al., submitted (see chapter 10.1.)].

Table 16: Replication testing for association with incident T2D in Europeans.

Relative risk per 1SD

CpG	Gene	LOLIPOP	KORA	Combined	p-value
cg19693031	<i>TXNIP</i>	0.80 (0.63 to 1.00)	0.61 (0.48 to 0.78)	0.70 (0.59 to 0.83)	2.5×10^{-5}
cg09152259	<i>PROC</i>	0.93 (0.78 to 1.11)	0.94 (0.77 to 1.15)	0.94 (0.82 to 1.07)	0.32
cg04999691	<i>C7orf29</i>	1.01 (0.88 to 1.16)	0.90 (0.73 to 1.12)	0.98 (0.87 to 1.10)	0.71
cg11024682	<i>SREBF1</i>	1.12 (0.97 to 1.29)	1.38 (1.10 to 1.73)	1.19 (1.05 to 1.34)	5.4×10^{-3}
cg02650017	<i>PHOSPHO1</i>	0.81 (0.69 to 0.94)	0.79 (0.61 to 1.02)	0.80 (0.70 to 0.92)	1.2×10^{-3}
cg18181703	<i>SOCS3</i>	0.89 (0.76 to 1.04)	0.66 (0.52 to 0.84)	0.81 (0.71 to 0.92)	1.6×10^{-3}
cg06500161	<i>ABCG1</i>	1.17 (0.99 to 1.38)	1.61 (1.28 to 2.03)	1.31 (1.14 to 1.49)	1.2×10^{-4}
Methylation Score	-	1.63 (1.27 to 2.09)	2.24 (1.70 to 2.96)	1.88 (1.56 to 2.26)	2.5×10^{-11}

Relative risk Q4 vs Q1

CpG	Gene	LOLIPOP	KORA	Combined	p-value
cg19693031	<i>TXNIP</i>	1.32 (0.83 to 2.11)	5.75 (1.99 to 16.63)	1.68 (1.09 to 2.58)	1.8×10^{-2}
cg09152259	<i>PROC</i>	1.19 (0.73 to 1.95)	1.50 (0.72 to 3.11)	1.28 (0.85 to 1.93)	0.23
cg04999691	<i>C7orf29</i>	1.00 (0.60 to 1.66)	1.25 (0.49 to 3.17)	1.05 (0.68 to 1.64)	0.82
cg11024682	<i>SREBF1</i>	1.54 (0.92 to 2.56)	1.57 (0.61 to 4.05)	1.54 (0.99 to 2.42)	5.8×10^{-2}
cg02650017	<i>PHOSPHO1</i>	2.16 (1.26 to 3.69)	1.50 (0.53 to 4.21)	2.00 (1.24 to 3.22)	4.3×10^{-3}
cg18181703	<i>SOCS3</i>	1.59 (0.99 to 2.56)	4.75 (1.62 to 13.96)	1.90 (1.23 to 2.94)	3.7×10^{-3}
cg06500161	<i>ABCG1</i>	1.13 (0.71 to 1.79)	4.00 (1.50 to 10.66)	1.42 (0.94 to 2.16)	9.9×10^{-2}
Methylation Score		2.16 (1.27 to 3.66)	20.00 (2.68 to 149.0)	2.49 (1.50 to 4.15)	4.6×10^{-4}

Results are presented as OR (95% CI) per 1 SD increase in methylation for Q4 compared to Q1, where Q1 is the quartile with the lowest T2D risk. P-value is shown for the combined analysis [modified after Chambers et al., submitted (see chapter 10.1.)].

Methylome-wide data of the discovery and the prospective replication studies were combined using inverse variance meta-analysis. All five markers in this study reached methylome-wide significance for association with T2D (p-values 1.4×10^{-9} to 1.2×10^{-17}). Between the top and bottom quartiles of DNA methylation the relative risk for incident T2D is 1.05 to 2.00 for the analysis of the Europeans (table 17). The population risk for T2D due to methylation score (above 25th centile) was 32% in Europeans and 44% in Indian Asians.

Table 17: Relative risk for incident T2D for Europeans (replication).

CpG	Gene	Europeans (replication)
cg19693031	<i>TXNIP</i>	1.68 (1.09 to 2.58)
cg09152259	<i>PROC</i>	1.28 (0.85 to 1.93)
cg04999691	<i>C7orf29</i>	1.05 (0.68 to 1.64)
cg11024682	<i>SREBF1</i>	1.54 (0.99 to 2.42)
cg02650017	<i>PHOSPHO1</i>	2.00 (1.24 to 3.22)
cg18181703	<i>SOCS</i>	1.90 (1.23 to 2.94)
cg06500161	<i>ABCG1</i>	1.42 (0.94 to 2.16)
Methylation Score	-	2.49 (1.50 to 4.15)

Results are expressed as OR for Q4 compared to Q1. Q1 is the quartile with the lowest T2D risk. [modified after Chambers et al., submitted (see chapter 10.1.)]

3.2. Association with measures of glucose metabolism

For the analysis of methylome-wide association with measures of glucose metabolism a cross-sectional study was designed including samples from the KORA follow-up study F4.

3.2.1. Study characteristics

Study characteristics for fasting glucose, 2-hour glucose, fasting insulin, and HbA1c are shown in table 18 as well as appendix table 4 for 2-hour insulin and appendix table 5 for HOMA-IR. The median age in the study population was 59 years and 47.1% were male. The individuals included in the study were tendentially overweight with a median BMI of 27.06 kg/m². The glyceemic parameters were in normal range concerning T2D. 77.28% are normal glucose tolerant, whereas impaired fasting glucose, impaired glucose tolerance, and a combination of both were assessed in 4.97%, 14.36%, and 3.38% of the study population, respectively. 60.15% were physically active (table 18).

Study characteristics of participants included in the methylome-wide association study with 2-hour insulin (appendix table 4) and HOMA-IR (appendix table 5) are comparable to those for fasting glucose, 2-hour glucose, fasting insulin, and HbA1c. For 2-hour insulin the median age, which is 68 years, differs slightly from the other study populations. Furthermore, it included more subjects with impaired glucose tolerance as well as current smokers (appendix table 4).

Table 18: Study characteristics of the study population for DNA methylation analysis (n=1,448) for fasting glucose, 2-hour glucose, fasting insulin, and HbA1c.

	Median (25th; 75th percentile)	%
Sex [% male]	-	47.1
Age [years]	59 (53; 67)	-
BMI [kg/m²]	27.06 (24.51; 30.05)	-
Waist circumference [cm]	93.50 (84.45; 102.33)	-
Fasting serum glucose [mmol/l]	5.28 (4.94; 5.61)	-
2-hour serum glucose [mmol/l]	6.00 (5.00; 7.17)	-
HbA1c [%]	5.50 (5.20; 5.70)	-
Glucose tolerance status [%]		
NGT	-	77.28
IFG	-	4.97
IGT	-	14.36
Combined IFG and IGT	-	3.38
Insulin [μU/ml]	4.10 (2.80; 6.70)	-
2-hour insulin [μU/ml]	50.20 (28.70; 412.00)	-
HOMA-IR	0.97 (0.64; 1.59)	-
C-reactive protein [mg/l]	1.15 (0.58; 2.23)	-
Leucocytes [/nl]	5.50 (4.70; 6.50)	-
Total cholesterol [mmol/l]	5.71 (5.11; 6.43)	-
Triglycerides [mmol/l]	1.22 (0.86; 1.71)	-
Systolic blood pressure [mmHg]	122.20 (111.00; 134.50)	-
Diastolic blood pressure [mmHg]	75.50 (69.50; 82.50)	-
Alcohol consumption [g/day]	8.43 (0.00; 22.86)	-
Smoking status [%]		
never	-	44.75
ex	-	40.12
current	-	15.12
Physically active [%] (combination of activity during summer and winter)	-	60.15
CD8⁺ T cells [%]⁺	0.09 (0.05; 0.14)	-
CD4⁺ T cells [%]⁺	0.16 (0.12; 0.21)	-
Natural killer cells [%]⁺	0.02 (0.00; 0.04)	-
B cells [%]⁺	0.05 (0.03; 0.06)	-
Monocytes [%]⁺	0.12 (0.10; 0.13)	-
Granulocytes [%]⁺	0.63 (0.57; 0.69)	-

+ Data are presented for the estimated white blood cell proportions using a recently published method [142]

NGT: normal glucose tolerance; IFG: impaired fasting glucose; IGT: impaired glucose tolerance

3.2.2. Genome-wide DNA methylation analysis

Results of the association analyses between the degree of DNA methylation and parameters of glucose metabolism are displayed in table 19. Furthermore, results for fasting glucose, 2-hour glucose, fasting insulin, and HOMA-IR for model 1 are also shown as Manhattan plots in figure 20. Annotation information of detected CpG sites is summarized in appendix table 6.

Table 19: Results for association of genome-wide DNA methylation levels and fasting glucose, 2-hour glucose, fasting insulin, and HOMA-IR as well as 2-hour insulin with reduced number of CpG sites after adjustment for different potential confounders.

Phenotype	CpG	Gene	Age, sex, estimated white blood cell proportions (model 1)			Age, sex, estimated white blood cell proportions, BMI (model 2)		
			Coefficient	p-value	B-H-adjusted p-value	Coefficient	p-value	B-H-adjusted p-value
Fasting glucose	cg00574958	<i>CPT1A</i>	-0.097	9.5x10⁻⁸	0.014	-0.060	5.8x10 ⁻⁴	0.971
	cg06500161	<i>ABCG1</i>	0.043	3.1x10⁻¹⁰	6.8x10⁻⁵	0.026	1.3x10 ⁻⁴	0.852
	cg07504977	unannotated	0.022	3.4x10⁻⁷	0.035	0.017	5.9x10 ⁻⁵	0.728
	cg11024682	<i>SREBF1</i>	0.056	3.1x10⁻¹⁰	6.8x10⁻⁵	0.041	1.9x10 ⁻⁶	0.592
	cg22040809 [#]	<i>HCG11</i>	-0.021	4.0x10⁻⁷	0.035	-0.017	9.9x10 ⁻⁶	0.611
2-hour glucose	cg06500161	<i>ABCG1</i>	0.043	3.0x10⁻⁹	1.3x10⁻³	0.027	1.7x10 ⁻⁴	0.999
Fasting insulin	cg06500161	<i>ABCG1</i>	0.045	1.8x10⁻⁹	8.1x10⁻⁴	0.019	0.007	0.385
	cg07092212	<i>DGKZ</i>	-0.076	7.4x10⁻⁷	0.041	-0.062	8.5x10 ⁻⁶	0.210
	cg09613192	unannotated	0.022	1.0x10⁻⁷	0.015	0.014	2.8x10 ⁻⁴	0.360
	cg09694782	unannotated	-0.037	5.5x10⁻⁸	0.012	-0.031	4.3x10 ⁻⁷	0.071
	cg11376147	<i>SLC43A1</i>	-0.075	2.4x10⁻⁷	0.021	-0.040	0.002	0.364
	cg17266233	<i>DGKZ</i>	-0.099	2.9x10⁻⁷	0.022	-0.086	1.0x10 ⁻⁶	0.071
	cg17971578	<i>STK40</i>	-0.038	3.8x10⁻⁷	0.024	-0.019	0.006	0.381
	cg22065733	unannotated	-0.065	1.6x10⁻⁷	0.017	-0.056	8.9x10 ⁻⁷	0.071
2-hour insulin*	cg00574958	<i>CPT1A</i>	-0.103	7.1x10⁻⁴	3.5x10⁻³	-0.064	0.027	0.079
	cg04161365	<i>DHRS13</i>	-0.029	0.016	0.030	-0.018	0.125	0.181
	cg06500161	<i>ABCG1</i>	0.059	1.6x10⁻⁷	2.4x10⁻⁶	0.040	1.7x10⁻⁴	2.6x10⁻³
	cg09694782	unannotated	-0.036	6.8x10⁻⁴	3.5x10⁻³	-0.031	1.6x10⁻³	0.012
	cg11024682	<i>SREBF1</i>	0.043	2.5x10⁻³	6.4x10⁻³	0.031	0.021	0.078
	cg11376147	<i>SLC43A1</i>	-0.074	1.2x10⁻³	3.6x10⁻³	-0.043	0.047	0.118
	cg13016916	<i>CREB3L2</i>	0.015	3.1x10⁻³	6.6x10⁻³	0.013	4.2x10⁻³	0.021
	cg17971578	<i>STK40</i>	-0.041	9.7x10⁻³	3.6x10⁻³	-0.020	0.101	0.181
	cg20477259	<i>TNF</i>	-0.040	0.021	0.035	-0.013	0.441	0.472
HOMA-IR**	cg04161365	<i>DHRS13</i>	-0.040	9.9x10⁻⁷	0.043	-0.028	1.3x10 ⁻⁴	0.266
	cg06500161	<i>ABCG1</i>	0.047	1.7x10⁻¹⁰	7.5x10⁻⁵	0.021	2.6x10 ⁻³	0.307
	cg07092212	<i>DGKZ</i>	-0.082	7.8x10⁻⁸	6.9x10⁻³	-0.067	1.0x10 ⁻⁶	0.051
	cg09613192	unannotated	0.021	2.2x10⁻⁷	0.014	0.013	5.8x10 ⁻⁴	0.285
	cg09694782	unannotated	-0.037	6.4x10⁻⁸	6.9x10⁻³	-0.031	5.3x10 ⁻⁷	0.051

cg11376147	<i>SLC43A1</i>	-0.078	6.3x10⁻⁸	6.9x10⁻³	-0.042	1.3x10 ⁻³	0.297
cg13016916	<i>CREB3L2</i>	0.016	7.4x10⁻⁷	0.037	0.015	2.9x10 ⁻⁷	0.051
cg17266233	<i>DGKZ</i>	-0.102	1.5x10⁻⁷	0.011	-0.087	5.1x10 ⁻⁷	0.051
cg17971578	<i>STK40</i>	-0.038	3.9x10⁻⁷	0.022	-0.018	7.7x10 ⁻³	0.338
cg20477259	<i>TNF</i>	-0.054	1.1x10⁻⁶	0.043	-0.030	3.4x10 ⁻³	0.310
cg22065733	unannotated	-0.067	7.1x10⁻⁸	6.9x10⁻³	-0.057	4.0x10 ⁻⁷	0.051

Genome-wide association of DNA methylation and T2D related traits were calculated with a linear mixed effects model.

CpG site has to be regard conditionally as it is listed as a cross reactive probe by Chen *et al.* [137]

* Analyses for 2-hour insulin were performed with a reduced data set (n=617) and for only CpG sites (n=15) which were significantly associated with the other investigated phenotypes

** Analyses for HOMA-IR were performed with a reduced data set (n=1,440)

Bold marks CpG sites which are significant after correction for Benjamini-Hochberg in the different adjustment models.

In total, methylation levels of 15 CpG sites showed genome-wide significant associations with measures of glucose metabolism. Methylation at five CpG sites showed genome-wide significant associations with fasting glucose in model 1 (B-H-adjusted p-values between 6.8×10^{-5} and 0.035). Additionally, methylation level at one CpG site was statistically significantly associated with 2-hour glucose after adjustment model 1 (B-H-adjusted p-value 1.3×10^{-3}). The methylation degrees of eight CpG sites were significantly associated with fasting insulin in model 1 with B-H-adjusted p-values between 8.1×10^{-4} and 0.041. For HOMA-IR, a genome-wide association (B-H adjusted p-values between 7.5×10^{-5} and 0.043) was detected after adjustment model 1 for methylation levels of eleven CpG sites. For HbA1c no methylome-wide significant association was found. Testing the association of all aforementioned CpG sites with 2-hour insulin, for nine CpG sites a significant association between DNA methylation and the phenotype was observed in model 1 (B-H-adjusted p-values between 2.4×10^{-6} and 0.35). Methylation degree at three CpG sites was still significantly associated with 2-hour insulin in model 2. Furthermore, evidence suggests associations of some CpG sites with fasting as well as 2-hour insulin and HOMA-IR, as they are borderline non-significant (B-H-adjusted p-values between 0.05 and 0.079). Comparing the beta-coefficients in model 1 vs model 2, on average 29.2% of the association with fasting glucose in model 1 could be explained by the BMI (range 19.0-39.5%). Furthermore, BMI explained 37.2% of the association between DNA methylation and 2-hour insulin, for fasting insulin on average 31.6% in model 1 (range 13.1-57.8%), for 2-hour insulin on average 35.9% in model 1 (range 13.3-67.5%), and for HOMA-IR on average 30.6% in model 1 (range 6.3-55.3%).

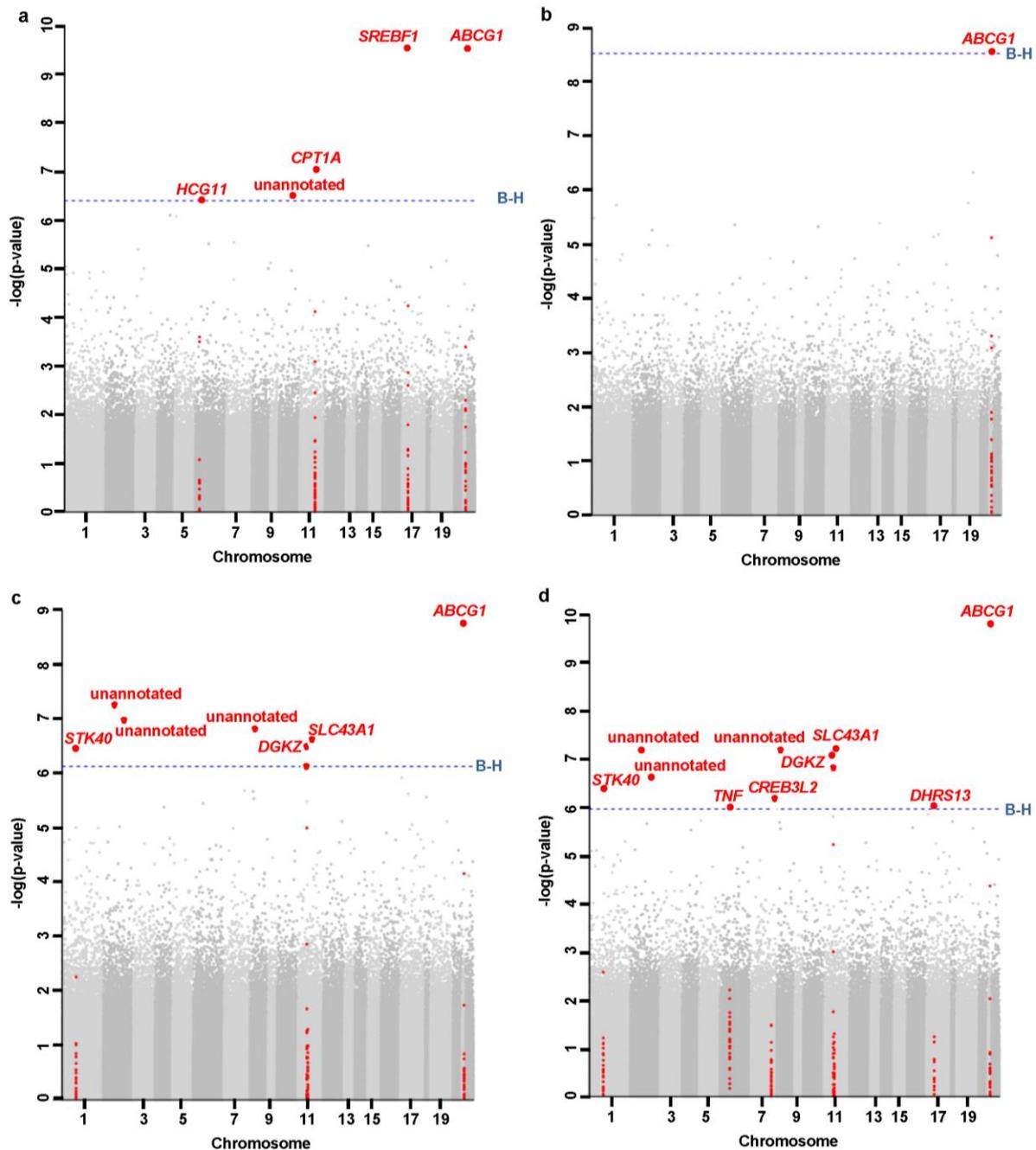


Figure 20: Genome-wide associations between methylation and (a) fasting glucose level, (b) 2-hour glucose level, (c) fasting insulin level, (d) HOMA-IR after adjustment for sex, age, and estimated white blood cell proportions (model 1). Results are plotted for each annotated chromosome (excluding the sex chromosomes) (x-axis) against the $-\log(p\text{-value})$ (y-axis). The Benjamini-Hochberg method was used for correction for multiple testing. Red dots mark significant loci. The dotted line marks the significance threshold.

3.2.3. DNA methylation quintile analysis

Trend analyses of methylation degree and selected phenotypes are performed to see an association between them. Analyzing the association between DNA methylation in quintiles at the three CpG sites [cg06500161 (*ABCG1*), cg09694782 (unannotated), and cg13016916 (*CREB3L2*)] which are still significant in model 2 for 2-hour insulin and other phenotypes, significant associations between methylation degree and some variables were detected, mostly for cg06500161. Here for BMI, waist circumference, fasting glucose, 2-hour glucose, triglycerides, and sex the most significant associations were found. Furthermore, CD8⁺ T cells, monocytes, and granulocytes (which are estimated [142]) were significantly associated with the methylation degree at cg06500161. For cg09694782 significant associations of the methylation level were detected with age, fasting insulin, and HOMA-IR, as well as all estimated white blood cell proportions [142]. For cg13016916, it was not possible to detect a significant association of the investigated phenotypes and methylation degree (table 20, appendix table 7). Results for BMI, fasting insulin, 2-hour insulin, and HOMA-IR are also presented in figure 21.

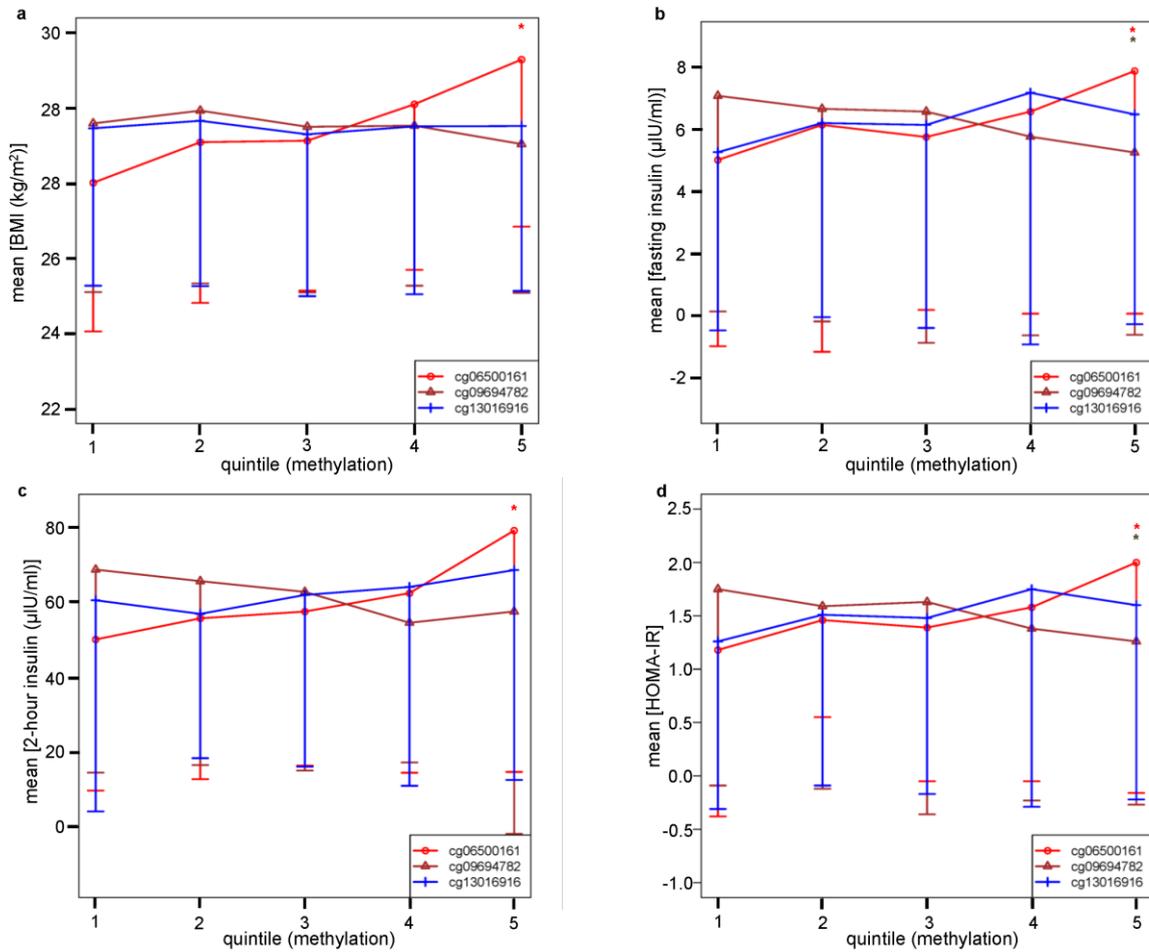


Figure 21: Distribution of BMI (a), fasting insulin (b), 2-hour insulin (c), and HOMA-IR (d) (y-axis) for the different degrees of methylation presented as quintiles (x-axis). * represents the corresponding CpG site, where the p for trend was significant after Bonferroni correction.

Table 20: Association between DNA methylation at cg06500161 and different phenotypes based on quintiles of methylation level.

	1.Quintile (n=290)	2.Quintile (n=289)	3. Quintile (n=290)	4. Quintile (n=289)	5.Quintile (n=290)	
Phenotype	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	p for trend (bonf. adjusted)
Age [years] #	58.22 (8.77)	58.95 (8.44)	60.50 (8.46)	60.13 (8.53)	61.47 (9.10)	2.90x10 ⁻⁵
BMI [kg/m ²] #	26.01 (3.96)	27.09 (4.28)	27.13 (3.99)	28.10 (4.41)	29.29 (4.45)	1.13x10 ⁻²⁰
Waist circumference [cm]	87.39 (12.33)	92.04 (12.21)	92.29 (11.27)	95.93 (12.53)	100.60 (12.42)	1.64x10 ⁻³⁸
Fasting glucose [mmol/l] #	5.12 (0.47)	5.24 (0.48)	5.32 (0.53)	5.36 (0.52)	5.52 (0.55)	6.54x10 ⁻²¹
2-hour glucose [mmol/l] #	5.71 (1.53)	6.16 (1.70)	6.15 (1.72)	6.28 (1.67)	6.80 (1.75)	7.94x10 ⁻¹³
HbA1c [%]	5.41 (0.33)	5.45 (0.32)	5.45 (0.29)	5.46 (0.32)	5.57 (0.31)	2.31x10 ⁻⁷
C-reactive protein [mg/l] #	1.68 (1.76)	1.61 (1.57)	1.70 (1.57)	1.80 (1.74)	1.84 (1.67)	1
Fasting insulin [μIU/ml] #	5.02 (6.01)	6.15 (7.32)	5.76(5.58)	6.58 (6.52)	7.89 (7.83)	6.26x10 ⁻⁶
2-hour insulin [μIU/ml] #	50.12 (40.46)	55.78 (43.08)	57.6 (41.29)	62.53 (48.11)	79.3 (64.67)	1.77x10 ⁻⁵
HOMA-IR #	1.18 (1.56)	1.46 (0.91)	1.39 (1.44)	1.58 (1.63)	2.00 (2.16)	3.44x10 ⁻⁷
Cholesterol [mmol/l] #	5.90 (0.97)	5.85 (0.97)	5.79 (0.99)	5.83 (1.02)	5.63 (1.06)	0.041
Triglycerides [mmol/l] #	1.09 (0.67)	1.25 (0.64)	1.39 (1.22)	1.59 (0.89)	1.91 (1.23)	1.06x10 ⁻²⁶
Systolic blood pressure [mm Hg]	119.94 (19.09)	121.52 (16.38)	125.01 (18.58)	123.09 (17.78)	126.97 (18.53)	4.39x10 ⁻⁵
Diastolic blood pressure [mm Hg]	75.14 (10.25)	75.81 (8.89)	76.77 (9.97)	75.84 (9.78)	77.5 (10.32)	0.113
CD8 ⁺ T cells #+	0.08 (0.05)	0.09 (0.06)	0.10 (0.07)	0.11 (0.07)	0.12 (0.07)	4.71x10 ⁻¹³
CD4 ⁺ T cells +	0.16 (0.06)	0.17 (0.06)	0.16 (0.06)	0.17 (0.06)	0.16 (0.06)	1
Natural killer cells #+	0.02 (0.02)	0.02 (0.02)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)	0.082
B cells #+	0.05 (0.04)	0.05 (0.03)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	1
Monocytes +	0.11 (0.02)	0.11 (0.02)	0.12 (0.02)	0.12 (0.02)	0.12 (0.03)	1.28x10 ⁻¹⁰
Granulocytes +	0.65 (0.09)	0.63 (0.09)	0.63 (0.09)	0.62 (0.09)	0.62 (0.09)	1.35x10 ⁻⁵

	Quintile (n=290)			2. Quintile (n=289)			3. Quintile (n=290)			4. Quintile (n=289)			5. Quintile (n=290)		
	number	number	p-value	Number	p-value	number	p-value	number	p-value	number	p-value	number	p-value		
Sex [male/female]	83/207	122/167	8.59x10 ⁻⁴ *	127/163	2.03x10 ⁻⁴ *	156/133	9.81x10 ⁻¹⁰ *	194/96	6.01x10 ⁻²⁰ *						
Glucose status [combination of IFG and IGT / IFG / IGT / normal]	3/8/21/258	8/6/46/229	1.99x10 ⁻³ *	11/13/40/226	3.50x10 ⁻³	6/23/46/214	5.00x10 ⁻⁴ *	21/22/55/192	5.00x10 ⁻⁴ *						

Means, standard deviations and p-values for trend (Bonferroni corrected) are presented for the different quintiles for the continuous phenotypes, and total numbers in the different quintiles and p-values (comparing the quintile vs 1st quintile) for categorical variables. * p-value was significant after correction for multiple testing. # variables were log transformed for determination of p-values); + cell types were estimated using method developed by Houseman *et al.* [142], IFG: impaired fasting glucose, IGT: impaired glucose tolerance

3.2.4. Pathway analysis

To put results into a functional context, pathway analysis was conducted using Ingenuity Software. “Phospholipase C Signaling” reached genome-wide significance after adjustment for age, sex, and estimated white blood cell proportions (model 1) for fasting glucose looking at the canonical pathways (B-H-adjusted p-value of 0.039) (table 21). 21 of 231 genes included in the pathway are amongst the top 1,000 hits from the genome-wide methylation analysis. This signaling pathway was also detected after additional adjustment for BMI (B-H-adjusted p-value 0.178). Besides this pathway, “PI3K Signaling in B Lymphocytes” and “Ephrin Receptor Signaling” belong to the top pathways for the adjustment model 1 for fasting glucose with B-H-adjusted p-values of 0.059 each. The former signaling pathway was also under the top pathway for the model after additional adjustment for BMI (B-H-adjusted p-value 0.059). Results for the top ten canonical pathways for 2-hour glucose, fasting insulin, and HOMA-IR are shown in appendix table 8. For 2-hour glucose, a total of eleven canonical pathways reach the B-H-corrected level of significance (p-values < 0.05). “Thyroid Hormone Metabolism II (via Conjugation and/or Degradation)” was significant after adjustment for age, sex, and estimated white blood cell proportions, as well as after additional adjustment for BMI (B-H-adjusted p-values = 0.035 and 0.018, respectively). 6 of 26 as well as 7 of 26 genes included in the pathway can be found amongst the top 1,000 hits. For fasting insulin and HOMA-IR, no significant canonical pathways were found after adjustment for B-H.

Table 21: Pathway analysis based on the top 1,000 CpG sites associated with fasting glucose using Ingenuity.

Adjusted for age, sex, and estimated white blood cell proportions (model 1)			Adjusted for age, sex, estimated white blood cell proportions, and BMI (model 2)		
Ingenuity Canonical Pathways	B-H-adjusted p-value	Ratio	Ingenuity Canonical Pathways	B-H-adjusted p-value	Ratio
<u>Phospholipase C Signaling</u>	0.039	21/231	Ephrin Receptor Signaling	0.059	17/172
Ephrin Receptor Signaling	0.059	16/172	AMPK Signaling	0.059	14/133
PI3K Signaling in B Lymphocytes	0.059	13/123	Leptin Signaling in Obesity	0.085	9/73
CCR3 Signaling in Eosinophils	0.069	12/113	Molecular Mechanisms of Cancer	0.085	25/359
Ephrin B Signaling	0.079	9/73	RhoGDI Signaling	0.085	15/172
Role of NFAT in Cardiac Hypertrophy	0.079	15/176	Ephrin A Signaling	0.085	7/48
Corticotropin Releasing Hormone Signaling	0.079	11/108	Axonal Guidance Signaling	0.085	28/425
Fcy Receptor-mediated Phagocytosis in Macrophages and Monocytes	0.079	10/93	PPAR α /RXR α Activation	0.122	14/165
Hepatic Fibrosis/Hepatic Stellate Cell Activation	0.079	16/196	Phospholipase C Signaling	0.178	17/231
Thrombin Signaling	0.106	15/187	Integrin Signaling	0.178	15/195

The p-value corrected using the Benjamini-Hochberg method for multiple testing is presented for each pathway, besides the ratio (number of genes uploaded in the software/total number of genes included in the pathway). Underlined pathways are significant after correction for multiple testing using Benjamini-Hochberg.

3.2.5. Gene expression analysis

In order to assess the functional impact of DNA methylation on the transcriptome, the association between the degree of methylation and transcript levels in 533 subjects with overlapping methylation and expression data was investigated. The study characteristics of the samples are shown in table 22.

Table 22: Characteristics of the study population for gene expression analysis (n=533).

	Median (25 th ; 75 th percentile)	%
Sex [% male]	-	47.65
Age [years]	68 (65; 72)	-
BMI [kg/m²]	27.80 (25.23; 30.49)	-
Waist circumference [cm]	95.80 (88.60; 103.70)	-
Fasting serum glucose [mmol/l]	5.39 (5.06; 5.72)	-
2-hour serum glucose [mmol/l]	6.44 (5.39; 7.72)	-
HbA1c [%]	5.50 (5.30; 5.70)	-
Glucose tolerance status [%]		
NGT	-	70.17
IFG	-	5.44
IGT	-	18.57
Combined IFG and IGT	-	5.82
Insulin [µIU/ml]	4.70 (3.30; 7.40)	-
2-hour insulin [µIU/ml]	49.80 (28.50; 77.30)	-
HOMA-IR	1.13 (0.77; 1.80)	-
C-reactive protein [mg/l]	1.34 (0.71; 2.48)	-
Leucocytes [/nl]	5.60 (4.70; 6.40)	-
Cholesterol [mmol/l]	5.81 (5.12; 6.51)	-
Triglycerides [mmol/l]	1.23 (0.90; 1.69)	-
Systolic blood pressure [mmHg]	126.5 (114.5; 138.5)	-
Diastolic blood pressure [mmHg]	74.50 (68.50; 81.50)	-
Alcohol consumption [g/day]	7.86 (0.00; 20.43)	-
Smoking status [%]		
never	-	53.28
ex	-	39.77
current	-	6.94
Physically active [%] (combination of activity during summer and winter)	-	55.91
CD8⁺ T cells [%]+	0.09 (0.05; 0.15)	-
CD4⁺ T cells [%]+	0.15 (0.11; 0.19)	-
Natural killer cells [%]+	0.02 (0.00; 0.04)	-
B cells [%]+	0.04 (0.03; 0.06)	-

Monocytes [%]+	0.12 (0.10; 0.14)	-
Granulocytes [%]+	0.64 (0.58; 0.70)	-

+ Data are presented for the estimated white blood cell proportion using a recently published method [142]

NGT: normal glucose tolerance; IFG: impaired fasting glucose; IGT: impaired glucose tolerance

The study population has a median age of 68. Gender distribution is nearly equal, as in other study populations. Also, other characteristics are quite comparable to the other study populations used for the cross-sectional study (table 18, appendix tables 4 and 5).

The methylation level of the CpG site cg06500161 (*ABCG1*) showed a significant association with *ABCG1* gene expression level (B-H-adjusted p-value = 1.1×10^{-9}) after adjustment for age, sex, BMI, estimated white blood cell proportions [142], technical variables, and plate (table 23). This association remained significant after additional adjustment for fasting glucose, 2-hour glucose, fasting insulin, 2-hour insulin, and HOMA-IR (B-H-adjusted p-values = 8.3×10^{-9} , 7.5×10^{-9} , 1.7×10^{-9} , 4.4×10^{-8} , 2.4×10^{-9} , respectively). The associations between gene expression levels and the degree of methylation at the other 14 significant CpG sites found for the methylation analysis were not significant (appendix table 9); the analyses of gene expression and phenotypes also revealed no significant associations (data not shown).

Table 23: Summary of the analysis for association between DNA methylation and gene expression, as well as between DNA methylation and phenotypes.

	Association of methylation at cg06500161 (<i>ABCG1</i>) with different phenotypes					Association of methylation at cg06500161 and gene expression level of a transcript annotated to <i>ABCG1</i> (ILMN_2329927)					
	fasting glucose	2-hour glucose	fasting insulin	2-hour insulin	HOMA-IR	no adj. for phenotype	adj. for fasting glucose	adj. for 2-hour glucose	adj. for fasting insulin	adj. for 2-hour insulin	adj. for HOMA-IR
Coefficient	0.043	0.018	0.045	0.059	0.047	-3.623	-3.495	-3.507	-3.617	-3.398	-3.594
p-value	3.1×10^{-10}	3.0×10^{-9}	1.8×10^{-9}	1.6×10^{-7}	1.7×10^{-10}	2.5×10^{-12}	1.8×10^{-11}	1.7×10^{-11}	3.8×10^{-12}	9.7×10^{-11}	5.4×10^{-12}
Adj.p-value	6.8×10^{-5}	1.3×10^{-3}	8.1×10^{-4}	2.4×10^{-6}	7.5×10^{-5}	1.1×10^{-9}	8.3×10^{-9}	7.5×10^{-9}	1.7×10^{-9}	4.4×10^{-8}	2.4×10^{-9}

4. Discussion

A number of studies have demonstrated an influence of the methylation degree at different genes (see chapter 1.3.3.) on the development of diabetes, but there is a lack of studies with whole blood samples from participants with European ancestry. Therefore the presented projects focus on the determination of genome-wide methylation degree pattern and incident T2D, as well as measures of glucose metabolism.

4.1. Association with incident T2D

4.1.1. Main findings

To my knowledge this is the first study to analyze methylome-wide association with incident T2D using whole blood samples of participants of European ancestry alone. The methylation degrees of six CpG sites showed significant associations with incident T2D in the discovery study. However, the associations attenuated after additional adjustment for BMI, indicating an important impact of body weight on the association between DNA methylation and incident T2D at these loci. None of the leading CpG sites could be replicated using an independent method in the replication study. However, the methylation levels of three CpG sites flanking the leading CpG site annotated to *AKT2* and one of the CpG sites annotated to *C14orf182* were found to be significantly associated with incident T2D in the replication stage, even after adjustment for BMI for one CpG site per amplicon. The general functions of proteins encoded by genes annotated to the methylome-wide significant CpG sites are summarized in table 24. By analyzing a potential accumulation of DNA methylation data in pathways, several pathways were observed which can be linked to diabetes. None of the CpG sites annotated to genes for which a link between DNA methylation and diabetes has been reported (chapter 1.3.4.) are significant after B-H correction in this analysis.

Table 24: General function of proteins encoded by annotated genes of genome-wide significant CpG sites associated with incident T2D in humans.

Gene	Function
CASZ1	<u>castor zinc finger 1</u> Encodes for a zinc finger transcription factor. The protein may function as a tumor suppressor. SNPs of this gene are associated with blood pressure variation.
TMEM57	<u>transmembrane protein 57</u> The gene plays a major role in trafficking and lipid metabolism. SNPs near this gene are associated with serum lipid levels or total cholesterol [144, 145].
VIM	<u>Vimentin</u> Encodes for a member of the intermediate filament family, which, together with microtubules and actin microfilaments, form the cytoskeleton. This protein is responsible for maintaining cell shape, integrity of the cytoplasm, and stabilizing cytoskeletal interactions. It is also involved in immune response, and controls the transport of low-density lipoprotein cholesterol from the lysosome to the site of esterification.
C14orf182	<u>chromosome 14 open reading frame 182</u> no known function
AKT2	<u>v-akt murine thymoma viral oncogene homolog 2</u> It is a putative oncogene and encodes a protein belonging to the subfamily of serine/threonine kinases containing SH2-like domains. It is overexpressed in some ovarian carcinoma cell lines and primary ovarian tumors. Furthermore, the overexpression contributes to the malignant phenotype of a subset of human ductal pancreatic cancers.

Source: www.ncbi.nlm.nih.gov/gene/ on December 2014 unless stated otherwise.

Furthermore, data from the nested case-control study were included for the replication investigating methylome-wide association with incident T2D in peripheral blood amongst Indian Asians and Europeans in the LOLIPOP study. Five of the seven genome-wide significant CpG sites from the discovery study were replicated combining data from Indian Asians and Europeans. These included the same CpG sites annotated to *SREBF1* and *ABCG1* which were found in the analysis of measures of glucose metabolism, discussed in more detail in chapter 4.3. The general functions of proteins encoded by genes annotated to the methylome-wide significant CpG sites (excluding *ABCG1* and *SREBF1*) are summarized in table 25. The different findings can be explained by the different numbers of samples

included in the studies. In the KORA study, 196 matched case-control pairs with European ancestry were included, whereas the LOLIPOP-study included 1,074 Indian Asians with incident T2D patients as cases and 1,590 controls from the same ancestry. Therefore, different ethnicities may explain the different findings. A further possible explanation could be the shorter follow-up time of 5.3 years in the KORA study compared to 7.6 years in the LOLIPOP study. Different matching criteria for the case-control pairs is another possibility. For the KORA study age \pm 2 years, sex, and observation time to diabetes diagnosis were defined as matching criteria, whereas for the LOLIPOP study age a 5 year interval and sex were used.

Table 25: General function of proteins encoded by annotated genes of replicated genome-wide significant CpG sites associated with incident T2D in Indian Asians and Europeans.

Gene	Function
<i>TXNIP</i>	<u>thioredoxin interacting protein</u> This gene plays an important role in mammalian cells as it inhibits thioredoxin under oxidative stress conditions. Inhibition of it leads to increased reactive oxygen species in immune cells or hematopoietic cells [146].
<i>PHOSPHO1</i>	<u>phosphatase, orphan 1</u> Encodes for an orphan phosphatase that belongs to the family of halo-acid dehalogenases. The gene is highly expressed in bone and matrix vesicles. Inhibition leads to a reduction in the ability of matrix vesicles to mineralize [147].
<i>SOCS3</i>	<u>suppressor of cytokine signaling 3</u> Encodes for a member of the STAT-induced STAT inhibitor family, also known as suppressor of cytokine signaling. These members are cytokine-inducible negative regulators of cytokine signaling. Its expression is induced by different cytokines. The encoded protein can bind to JAK2 kinase, and inhibit JAK2 activity.

ABCG1 and *SREBF1* are discussed in chapter 4.3.1. Source: www.ncbi.nlm.nih.gov/gene/ on December 2014, unless stated otherwise.

4.1.2. Involvement of *AKT2* in diabetes

CpG sites annotated to the *AKT2* (v-akt murine thymoma viral oncogene homolog 2) gene, other than *C14o4f182*, were confirmatively identified to be associated with T2D in this study. Surprisingly, the methylation level of one CpG site annotated to *AKT2*, which is also covered by the Infinium HumanMethylation450 BeadChip (cg030239529), showed significant association with incident T2D in the replication study in an unadjusted model. It seems unlikely that this finding is caused by a lack of power, because the number of included samples in the replication study is about 2 times higher compared to the discovery study, where the uncorrected p-value was 0.54 for this CpG site. The different findings may be more attributable to the different methylation degree analysis.

From a biological point of view, *AKT2* seems a plausible candidate for incident T2D, as animal as well as human studies demonstrate an association of this gene with diabetes and related traits. In cultured cells siRNA mediated knockdown of *Akt2* leads to a significantly reduced insulin-stimulated glucose transport and *Glut4* translocation [148]. This is in line with two recent studies. Takenaka *et al.* demonstrated that FLJ00068, a guanine nucleotide exchange factor, regulates *Rac1* downstream of *Akt2*, which is followed by stimulation of glucose uptake in skeletal muscle from mice and cultured myocytes. Therefore, the knockdown leads to significantly decreased *Akt2* triggered *Glut4* translocation [149]. Furthermore, Ng *et al.* showed that *Akt2* is needed to introduce *Glut4* translocation in 3T3-L1 adipocytes (derived from mice) [150]. Also, studies on muscle tissue from wild type and *Akt2*-null mice allow us to assume that *Akt2* is essential for the full effect of brief calorie restriction on insulin-stimulated glucose uptake using physiological insulin [151]. In *Pten*-haplodeficient (*Pten*^{+/-}/*Akt2*^{+/+}) and *Pten*^{+/+}/*Akt2*^{+/+} mice, the activity of *Akt2* in skeletal muscle influences lipid accumulation in the liver [152]. Furthermore, insulin resistance in *Akt2* deficient mice was inhibited by an additional haplodeficiency of *Pten* [153]. This is in line with a study by Leavens *et al.* demonstrating that *Akt2* is an important component of the insulin dependent regulation of the lipid metabolism during insulin resistance in mice [154]. In another mouse study, an influence of *Akt2* on insulin resistance and elevated plasma triglycerides was detected, as *Akt2* deficient mice showed these phenotypes as well

as fed and fasting hyperglycemia, hyperinsulinemia, and glucose tolerance [155]. Furthermore, it was demonstrated that *Rac1* and *Akt* regulate insulin-stimulated glucose uptake via different parallel pathways which are dysfunctional in insulin resistant muscle in mice [156]. These findings are in line with a study by Lu *et al.* showing that deletion of *Akt1* and *Akt2* in livers from mice led to glucose intolerance and insulin resistance [157]. In mice liver, an immediate action of insulin on hepatic glucose production functions dependent on redirection of glucose-6-phosphate to glycogen due to *Akt2* was shown [158]. In contrast, renal improvement due to zinc in diabetic mice is associated with glucose metabolism signaling mediated by metallothionein and *Akt1*, but not *Akt2* [159]. Cho *et al.* observed a diabetes-like phenotype in *Akt2* knockout mice [160]. In rats and 3T3-L1 adipocytes, an increased expression of Ser/Thr kinase *Akt2* led to an improvement of insulin sensitivity [161]. For pigs, an association of SNPs of *Akt2* on diabetes was observed [162].

The literature also provides evidence linking *AKT2* and diabetes in humans. The knockdown of *AKT2* in human SGBS adipocytes leads to an almost complete inhibition of preadipocyte proliferation with effects on insulin-stimulated lipogenesis and anti-lipolytic effects of insulin [163]. In vitro, the *AKT2* expression was impaired in obese adipose tissue. Furthermore, the expression was inversely correlated with BMI and HOMA-IR [164]. Tan *et al.* found that a heterozygous loss-of-function mutation in *AKT2* can lead to severe insulin resistance and lipodystrophy in humans. However, they speculate that mutations in and close to *AKT2* are unlikely to contribute significantly to the risk of T2D [165]. Human muscle shows a lower protein content of *AKT2* in type I fibers compared to type II fibers. Furthermore, the phosphorylation-response to insulin was reduced for *AKT* and the insulin signaling was decreased in muscle in T2D patients compared to lean and obese subjects [166]. This is in line with a study by Vind *et al.* also demonstrating impaired insulin-stimulated *AKT2* activity [167]. Based on their study on skeletal muscle in lean and obese insulin-resistant humans Brozinick *et al.* assume that the ability of insulin to activate *AKT2* and *AKT3* may explain the reduced insulin-stimulated glucose transport in insulin resistance [168]. Furthermore, *AKT2* phosphorylation due to acute induction of muscle insulin resistance in humans was demonstrated [169]. Analyzing muscle tissue in 184 non-diabetic twins with kinase assays and

phosphor-specific western blots, an association between *AKT* activity and in vivo insulin sensitivity was observed. Therefore, it is suggested that *AKT* may control in vivo insulin resistance and potentially T2D [170]. Analyzing adipose tissue of insulin-resistant obese subjects showed that *AKT2* was increased in glucose transport [171]. In addition, Georg *et al.* describe a family with severe insulin resistance and diabetes due to a mutation of the *AKT2* gene [172]. In a study conducted in non-diabetic twins, Friedrichsen *et al.* described that insulin-stimulated glycogen synthase activity was positively associated with *AKT2* activity. Glycogen synthesis in muscle is reduced in T2D [173]. Furthermore, *AKT2* was assumed to be causal for T2D [44]. In contrast, Sun *et al.* wrote that *AKT2* is not a major cause of diabetes in a non-obese Chinese Han population characterized by insulin resistance [174]. The results of this study add to the enormous body of evidence regarding the implication of *AKT2* in glucose metabolism and suggest that the association may at least partly be influenced by epigenetic modifications.

4.1.3. Involvement of remaining associated CpG sites in diabetes

For *TMEM57* (transmembrane protein 57) an association with total cholesterol was found, which is a biomarker for diabetes [20], in a meta-analysis of 16 GWAS on lipid traits [145]. Furthermore, for *VIM* (vimentin) an up-regulation was found in Zucker fatty compared to Zucker lean rats. Also a significantly increased expression level of this gene was detected in Zucker fatty rats, afterwards returning to the level of Zucker lean rats, which is comparable to that of Zucker diabetic fatty rats [175]. For *CASZ1* (castor zinc finger 1) and *C14orf182* (chromosome 14 open reading frame 182) this study presents first evidence of their roles in glucose metabolism. Therefore, further studies are warranted to validate these findings.

4.1.4. Accumulation of DNA methylation data in different pathways and link to diabetes

In order to functionally integrate the results, a pathway analysis based on the 1,000 top hits for both adjustment models was conducted. For some of these pathways, a well-established link to diabetes can be found in the literature. For “NGF (nerve growth factor) Signaling” a study showed that electroacupuncture represents a supportive tool to control diabetic polyneuropathy development by modulating this pathway in diabetes-targeted neurons in rats [176]. Nakagaki *et al.* showed that epalrestat is a potential therapy for improving wound healing for diabetic people, involving the up regulation of *NGF* [177]. In studies in type 1 diabetes rats it was shown that NGF levels in serum and expression in the kidney were increased [178], and that neurotrophic and metabotropic potential of *Ngf* and another protein may be involved in the molecular mechanisms of stress and diabetes [179]. In cell culture experiments Pierucci *et al.* show that β -cell transcription and translation independent apoptosis are induced by *NGF* deletion [180]. Furthermore, “Integrin Signaling” could be linked to diabetes. Park *et al.* found that high glucose levels changed integrin expression patterns, which increases integrin $\alpha 3$ and $\beta 1$ in pericytes in mice [181]. Additionally, it was observed in mice that $\alpha 4$ integrins are involved in the development of high fat diet induced insulin resistance. The authors conclude that blocking of $\alpha 4$ integrin signaling can prevent the development of obesity-induced insulin resistance [182]. Human embryonic stem cell gene expression studies show an involvement of integrins and catenins in β -cell differentiation [183]. Furthermore, $\alpha 3$ and $\beta 1$ integrin-extracellular matrix interactions are important for the regulation of β -cell survival in human fetal or adult islet cells [184]. With respect to “G-protein gamma/beta Signaling”, based on their studies analyzing skeletal muscle of rats and mice, Osorio *et al.* speculate that ATP signals through the $\beta\gamma$ subunit of the G-protein activate *PI3K- γ* , *AKT*, *AS160/RAB8A* and therefore promote GLUT4 exocytosis and reduce GLUT4 endocytosis [185].

4.1.5. Involvement of genes found in Indian Asians and Europeans in diabetes

The methylation data of the nested case-control study were included in order to replicate findings in the analysis performed by the LOLIPOP study. *TXNIP* (thioredoxin interacting protein) expression is increased due to glucose [186]. Furthermore, in mice studies it was shown that β -cell specific *Txnip* deletion improves β -cell mass and protects against diabetes [187]. This is in line with findings by Xu *et al.* analyzing mice and primary human islets, demonstrating that β -cell *Txnip* is upregulated in diabetes, whereas *Txnip* deficiency protects against diabetes by preventing β -cell apoptosis [188]. Also in a mice study it was shown that an overabundance of *Txnip* resulted in impaired glucose, insulin, and pyruvate tolerance by upregulating *G6pc* through interaction with small heterodimer partners [189]. In cell cultured experiments a decreased *Txnip* mRNA and protein expression due to metformin exposure was observed [190]. For *SOCS3* (suppressor of cytokine signaling 3) an inactivation in leptin receptor-expressing cells of mice was shown to protect against diet-induced insulin resistance but not against obesity [191]. This was also demonstrated by other studies, for example one showing the importance of *Socs3* for inhibiting insulin resistance in the skeletal muscle of mice [192], and another revealing that overexpression of this gene in adipose tissue of mice leads to local but not systematic insulin resistance [193]. Furthermore, it was shown that β -cell specific *Socs3*-deficient mice were protected against the development of diabetes induced by streptozotocin [194]. For humans, it was shown that there was no change in *SOCS3* levels in muscle when exposing men with T2D to IL-6 compared to the placebo group [195]. Li *et al.* found that SNPs annotated to *SOCS3* were significantly associated with BMI in humans [196]. In contrast, it was found that there is no strong effect of common SNPs within the *SOCS3* gene on the development of T2D in humans [197]. For *PHOSPHO1* (phosphatase, orphan 1) this study presents first evidence of its role in glucose metabolism. Therefore, further studies are warranted to validate these findings.

ABCG1 and *SREBF1*, which were also detected in the analysis with measures of glucose metabolism, are discussed in more detail under chapter 4.3.1. These

findings are in line with results of the presented study and support the hypothesis that DNA methylation is involved in the regulation of incident T2D.

4.2. Association with measures of glucose metabolism

4.2.1. Main findings

Until now no study has been published analyzing the associations between methylation patterns in whole blood samples and measures of glucose metabolism in a large population-based study on a genome-wide scale. Altogether, the degree of methylation at 441,552 CpG sites in 1,448 subjects (and 617 and 1,440 for 2-hour insulin and HOMA-IR, respectively) was investigated and 15 genome-wide significant associations with fasting glucose, 2-hour glucose, fasting insulin, and HOMA-IR was observed. Associations were independent of age, sex, and estimated white blood cell proportions. For three CpG sites in model 2 the association with for 2-hour insulin was still significant after adjustment for BMI. Furthermore, evidence suggesting associations with other phenotypes after adjustment for BMI was found, which explained an average of around 30% of the association. cg06500161 (annotated to *ABCG1*) was associated with all phenotypes investigated (except for HbA1c). The general functions of all proteins encoded by detected genes are summarized in table 26. Furthermore, by trend analysis associations between the degree of DNA methylation and different phenotypes were found, primarily for cg06500161, but also for cg06500161 and cg09694782 for estimated white blood cell proportions [142].

Table 26: General function of proteins encoded by annotated genes of genome-wide significant CpG sites associated with glycemic parameters in humans.

Gene	Function
<i>ABCG1</i>	<u>ATP-binding cassette, sub-family G (WHITE), member 1</u> Encodes for a protein belong to the superfamily of ATP-binding cassette (ABC) transporters. The proteins transport different molecules across extra- and intra-cellular membranes. It is involved in macrophage cholesterol and phospholipids transport and is assumed to regulate cellular lipid homeostasis on other cell types.
<i>CPT1A</i>	<u>carnitine palmitoyltransferase 1a</u> By the sequential action of carnitine palmitoyltransferase I and carnitine palmitoyltransferase II, together with a carnitine-acylcarnitine

translocase, the mitochondrial oxidation of long-chain fatty acids is initiated. CPT I is the key enzyme in carnitine-dependent transport across the mitochondrial inner membrane. Deficiency of CPT I leads to a reduced rate of fatty acid beta-oxidation.

CREB3L2 cAMP responsive element binding protein 3-like 2
Encodes for a member of the oas1 bZIP transcription factor family. Members can dimerize but as homodimers only. The encoded protein is a transcriptional activator.

DHRS13 dehydrogenase/reductase (SDR family) member 13
no known function

DKGZ diacylglycerol kinase zeta
Encodes for proteins belonging to the eukaryotic diacylglycerol kinase family. It can reduce protein kinase C activity by regulating diacylglycerol levels in intracellular signaling cascade and signal transduction.

HCG11 HLA complex group 11 (non-protein coding)
no known function

SLC43A1 solute carrier family 43 (amino acid system L transporter) member 1
It belongs to the system L family of plasma membrane carrier proteins that transport large neutral amino acids.

SREBF1 sterol regulatory element binding transcription factor 1
Encodes for a transcription factor binding to sterol regulatory element-1, a decamer flanking the low density lipoprotein receptor gene. Some genes are involved in sterol biosynthesis. It is synthesized as a precursor attached to the nuclear membrane and endoplasmic reticulum.

STK40 serine/threonine kinase 40
Belongs to the serine/threonine kinase family and is essential in diverse signaling pathways associated with a wide range of cellular activities, including proliferation, differentiation, survival, and apoptosis [198]. It was shown to induce extra-embryonic endoderm differentiation from mouse embryonic stem cells [199].

TNF tumor necrosis factor
Encodes for a multifunctional proinflammatory cytokine belonging to the tumor necrosis factor superfamily. It is mainly secreted by macrophages. The cytokine is involved in the regulation of different biological processes including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation. Further, it is implicated in diseases like autoimmune diseases, insulin resistance, and cancer.

Source: www.ncbi.nlm.nih.gov/gene/ on December 2014 unless stated otherwise.

By implementing a second step an association between DNA methylation and gene expression of the *ABCG1* gene was demonstrated. *ABCG1* is an important regulator of cholesterol efflux from macrophages to HDL with a potential additional role in inflammatory signaling via TLRs [200]. By analyzing a potential accumulation of

DNA methylation data, several pathways were observed which can be linked to diabetes.

Finally, checking CpG sites annotated to the genes, an association between DNA methylation and diabetes were detected (chapter 1.3.4.); none of these are significant after B-H correction for the different adjustment models and phenotypes.

4.2.2. Involvement of *ABCG1* in diabetes

The findings support previous evidence by Hidalgo *et al.* who found an association between cg06500161 [*ABCG1* (ATP-binding cassette, sub-family G (WHITE), member 1)], fasting insulin and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study in CD4+ T cells [201]. Additionally, our study provides evidence of an association of cg06500161 with fasting glucose and 2-hour glucose (the latter was not analyzed in the study by Hidalgo *et al.*). Furthermore, new evidence of an implication of several other CpG sites in glucose homeostasis was presented. Reasons for the partly differing findings between the two studies can be (i) the DNA source (in this study, whole blood), (ii) the different normalization methods for methylation data, and (iii) different methods for adjustment or correction for multiple testing. Hidalgo and colleagues didn't have an adjustment for BMI. Based on their findings they conclude for a potential role of lipid metabolism in insulin resistance mediated via the ABC transporter. Furthermore, they speculate that not only the variation in methylation of *ABCG1* but also the underlying SNP variations mediate a part of the effects on insulin and HOMA-IR [201]. In the presented study, the associations for all investigated phenotypes, excluded 2-hour insulin, attenuated when adjusting for BMI. This is in line with well-established evidence of a close association between lipid levels and T2D, as for example shown in more than 5,000 men and women participating in the Framingham Heart Study [202]. In contrast to the Framing Heart Study, there has been no evidence that genetic variants of the *ABCG1* gene are associated with T2D in the general population, a lack emphasized in a study combining the Copenhagen General Population and Copenhagen City Hearty study with more than 40,000 participants in total [203].

Low HDL (high density lipoprotein) is known as an independent risk factor for T2D [204]. However, the connection between *ABCG1* and T2D as well as dyslipidemia is further supported by animal and gene expression studies. It has been shown in mice that a deficit in *Abcg1* as well as *Abca1* has a synergistic influence on β -cell function [205]. In another study, loss of *Abcg1* in pancreatic β -cells was shown to result in impaired insulin secretion and glucose tolerance, but without affecting cellular cholesterol content or efflux. Furthermore, *Abcg1* expression in β -cells isolated from diabetic mice may be influenced by thiazolidinediones [206]. These findings are in line with those derived from diabetic humans, as it was shown for patients with insulinomas that *ABCG1* expression in the tumor tissue was significantly associated with fasting insulin and insulin release index [207]. In addition, Johansson *et al.* showed that *ABCG1* is one of the most differentially expressed genes during weight loss and weight maintenance in adipose tissue from obese participants and at the three assessments (baseline, weight loss, and maintenance of reduced weight) these expression levels were correlated to predominant HDL concentrations [208]. Finally, Mauldin *et al.* found a 30% decrease in cholesterol efflux with a corresponding 60% improvement in cholesterol accumulation in macrophages of T2D patients in comparison to controls. Furthermore, *ABCG1* was not detectable in macrophages from T2D patients whereas it was detected in controls. The expression of *ABCG1* can be induced in both groups by treating them with liver x receptor agonist TO 901317 [209]. Taken together, the presented findings strongly support a role of the *ABCG1* gene in glucose metabolism and suggest that the association may in fact be mediated by epigenetic mechanisms.

Because it is known that DNA methylation can influence the expression of genes [59, 63, 71, 111, 115] the association of DNA methylation and gene expression was investigated in a next step. Here, an association between the degree of DNA methylation and gene expression level for cg06500161 (*ABCG1*) and *ABCG1* was found. The association was independent of fasting glucose, 2-hour glucose, fasting insulin, 2-hour insulin, and HOMA-IR. Therefore, it can be concluded that DNA methylation and gene expression influence each other and therefore have an effect

on insulin and glucose levels. The presented results do not provide any clues regarding the direction of influence.

4.2.3. Involvement of remaining associated CpG sites in diabetes

For *SREBF1* (sterol regulatory element binding transcription factor 1) a link can be found for diabetes or related traits. *SREBF1* is activated by insulin and may also influence dyslipidaemia and hepatic steatosis found in obesity, insulin resistance, and T2D [210]. During states with lower insulin levels, like fasting, *SREBF1* is decreased but, in contrast, during food uptake, obesity and insulin resistance, it rises [211]. Furthermore, Saxena *et al.* were able to demonstrate an association of a SNP annotated to *SREBF1* with T2D when analyzing participants with European ancestry [45]. For *CREB3L2* (cAMP responsive element binding protein 3-like 2), also significant in model 2 for 2-hour insulin, this study presents first evidence of its role in glucose metabolism. Therefore, further studies are warranted to validate these findings.

BMI could explain on average around 30% of the association of investigated phenotypes and detected CpG sites. The literature hints that BMI influences the DNA methylation pattern. For example Na *et al.* showed a differential influence of BMI on global DNA methylation in healthy women. They found a u-shape association between BMI and Alu methylation, where the lowest methylation degree was found at a BMI between 23 and 30kg/m². Due to these findings an involvement of BMI-related changes in Alu methylation in the etiology and pathogenesis of obesity is assumed [212]. Furthermore, an allele-specific, age-dependent, and BMI-associated methylation at *MCHR1* was shown in human blood samples [213]. Additionally, in this study, it was demonstrated by trend analysis that the degree of methylation at cg06500161 was rising with increasing measures of BMI. However, functional and/or time-series studies are warranted to elucidate potential cause and effect mechanisms in the association between those genes, adiposity, and glucose mechanisms. Due to the findings mentioned before, it can be assumed that BMI influences the methylation degree of CpG sites and therefore the association with measures of glucose metabolism is BMI-mediated. According to deductions, it is not

surprising that only some CpG sites are significant or rather borderline significantly associated with investigated phenotypes after adjustment for BMI. Some CpG sites where an association seems to be mediated by BMI are biologically plausible in the context of T2D or related traits.

Furthermore, for *CPT1A* (carnitine palmitoyltransferase 1a), which influences the rate limiting step of fatty acid transport into mitochondria for β -oxidation [214, 215], a link to diabetes or related traits can be found. For example Meng and colleagues could demonstrate in mouse pancreatic islets or rodent cell lines, that TO901317 influenced the lipotoxicity through decreasing the gene expression of *Cpt1a*, in addition to other genes, and therefore reducing β -cell function [215]. Nyman *et al.* could also show that a *Cpt1a* deficiency increases insulin sensitivity in mouse models exposed to either long term feeding or diet. Additionally, *Cpt1a*-deficient mice present different phenotypes depending on diet, demonstrating that diet, as well as genetics, influences the development of reduced glucose tolerance [216]. From a mice study, Orellana-Gavalda assumed that human-safe non-immunoreactive adeno-associated virus mediated *Cpt1a* expression could be used as molecular therapy for obesity and diabetes [217]. In contrast, Hirota *et al.* analyzed SNPs at *CPT1A* locus in diabetic against non-diabetic individuals and found that none of them were associated with T2D, hepatic lipid content or insulin resistance in T2D [218].

For *TNF* (tumor necrosis factor) a link to diabetes and related traits was found in the literature. This gene encodes a proinflammatory cytokine that is secreted mainly by macrophages and overexpressed in adipose tissue [219-221]. Besides others genes, it is involved in lipid metabolisms and implicated in insulin resistance [222-224]. Xu *et al.* demonstrated that $TNF\alpha$ concentration is significantly higher in patients with gestational diabetes compared to controls, independent of BMI [225]. This is in agreement with Chen *et al.* finding the same results in rats with T2D compared to non-diabetic controls [226] as well as Volpe *et al.* analyzing plasma from T2D patients and supernatant from palmitate-stimulated PBMCs [227]. Neutralization of $TNF\alpha$ signaling can improve insulin sensitivity [219, 222].

For *HCG11* [HLA complex group 11 (non-protein coding)], *DGKZ* (diacylglycerol kinase zeta), *SLC43A1* [solute carrier family 43 (amino acid system L transporter) member 1], *STK40* (serine/threonine kinase 40), and *DHRS13* [dehydrogenase/reductase (SDR family) member 13], this study presents first evidence of their roles in glucose metabolism. Therefore, further studies are warranted to validate these findings. Results for cg22040809 (*HCG11*), significant in adjustment model 1, have been regarded conditionally as it is listed by Chen *et al.* [137] as a cross reactive probe with other CpG sites. Therefore, the interference of another CpG site cannot be excluded.

4.2.4. Accumulation of DNA methylation data in different pathways and link to diabetes

In order to functionally integrate the results, a pathway analysis based on the 1,000 top hits for each phenotype and adjustment models was conducted. Some of the identified pathways have a well-established link to diabetes or related traits. Examples are “Leptin Signaling in Obesity”, which is involved in the regulation of obesity, an important risk factor for T2D [228], and “Ephrin A/B Signaling”. For the latter it was shown that EphA and ephrin A regulate insulin secretion [229] and that the communication of EPH receptor and ephrin between exocrine and endocrine cells is involved in pancreatic function [230]. Ephs and ephrins are expressed in pancreatic β -cells in humans and mice [231] and EphAs are tyrosine phosphorylated under low glucose concentrations and initiate forward signaling, which in turn reduces insulin secretion [232]. Furthermore, “Netrin Signaling” is a promising pathway in connection to diabetes, as in human and mice studies netrin 1 was detected in obese but not lean adipose tissue. Also it was shown that a hematopoietic deletion of *Ntn1* improves insulin sensitivity amongst other phenotypes [233]. Also Jayakumar and colleagues show that urinary netrin 1 is significantly increased in normoalbuminuric diabetes patients by comparing them with healthy patients and conclude that it can be a potential biomarker for predicting the development of chronic kidney disease in diabetes patients [234]. Additionally, for “Phospholipase C Signaling” a link to diabetes and related traits can be found. For phospholipase C β 3, a strong down-regulation in diabetic peripheral neuropathy

of diabetic mice was shown [235]. Furthermore, phospholipase C $\delta 1$ is involved in obesity by regulating thermogenesis and adipogenesis in mice [236] and has an effect on insulin secretion in a pancreatic β -cell line [237].

4.3. General challenges

Different challenges, such as for instance problems with the stability of induced changes, need to be faced for epigenetic analysis. Another point which has to be considered regards cell type-dependent epigenetic mechanisms. The degree of methylation is different across different cells and tissues [59, 62, 142]. This has to be kept in mind for both heterogeneous tissues and whole blood. Therefore, adjustment for the cell composition is necessary. Particularly, epigenetic patterns of disease-relevant tissues might further advance our understanding of disease pathophysiology. However, for ethical and practical reasons, DNA methylation analyses are often only feasible in whole blood rather than in disease-relevant tissues, particularly in large population-based observational studies. As no directly measured cell composition was available for the study samples of this thesis, estimated white blood cell proportions using the method developed by Houseman *et al.* [142] were used. In addition, it is often difficult to conclude if a change in DNA methylation status is the cause or result of a disease, leading to a discussion comparable to the chicken and egg dilemma [62, 101].

Furthermore, bioinformatical as well as biostatistical methods have to be developed to analyze the high-throughput data. For example the shift of the two chemical technologies used (Infinium I and Infinium II) for the Infinium HumanMethylation450 BeadChip has to be considered. Batch effects due to different protocols or handling have to be solved as well.

Another important point is the consideration of potential T2D risk factors/variables which could be potential confounders for the investigated phenotype. Besides age and sex, BMI is an important risk factor for T2D and related traits. In the presented studies many models were adjusted for BMI in light of this fact. Another possibility

could also be a stratification according BMI. This wasn't conducted for reasons of limited power.

Recent evidence showed that 5-hydroxymethylation is also an important regulator in the development of diseases. At present many available methods cannot distinguish between 5-methylcytosine and 5-hydroxymethylation [238, 239], but currently several approaches which will enable this discrimination are being developed [239, 240]. Booth and colleagues for example used oxidative bisulfite sequencing in combination with representation bisulfite sequencing (RRBS) [241] to distinguish between the kinds of methylation.

4.4. Strengths and limitations

One of the strengths of this thesis is the genome-wide approach, and the combination of different methods, like DNA methylation and gene expression as well as pathway analysis, to obtain insight into potential mechanisms behind glucose and insulin regulation. The quality of the Infinium HumanMethylation450 BeadChip as well as EpiTYPER[®] data was controlled during the laboratory and statistical process. All statistical models were adjusted for different potential confounding factors. Due to the study design of the nested case-control study it can be assumed that aberrant DNA methylation of the identified CpG sites increases the risk for T2D. In contrast, a major caveat of the findings of the cross-sectional study is that it does not allow deductions on cause and effect. A further strength of the presented studies is the population-based design, which included also 2-hour glucose and 2-hour insulin from OGTT. As it is known that DNA methylation patterns differ in different tissues [59, 62], the use of whole blood can be seen as a limitation for the presented studies, because it represents a mixture of cell types and may differ in methylation profiles from insulin-responsive tissues such as liver, skeletal muscle, and adipose tissue. Another limitation is the unavailability of direct measured cell proportions. However, an adjustment for estimated white blood cell proportions based on the methylation data for blood cells was performed to reduce potential confounding simply due to interindividual differences in blood cell proportions. In addition, whole blood is also a strength because it has the potential to be used for

biomarker studies. Further, peripheral blood is an important biological material in the context of prediction of diabetes as the samples are easily accessible in the clinical routine compared to tissue biopsies from the aforementioned organs. Other studies have also shown that it is a good surrogate for more disease-targeted tissues and cells in some cases. A further limitation of the two presented studies is that they did not distinguish between 5`methylcytosine and 5`hydroxymethylation, which was recently proven to have an effect on diseases. This is the largest study with participants of European ancestry so far and it had the potential to identify several epigenetic modifications related to T2D and glucose metabolism. Comparison of findings from the KORA- and LOLIPOP studies shows how study design and the number of samples can influence results. Due to small effect sizes, meta-analysis with large sample sizes seems promising to increase the statistical power. Last but not least, the biological relevance of many results of the presented studies were demonstrated.

4.5. Conclusion

At present there is no study published analyzing the association of DNA methylation patterns and T2D or measures of glucose metabolism in whole blood samples in a genome-wide approach. For the nested case-control study, analyzing the association with incident T2D, the methylation level of six CpG sites reach genome-wide significance independent of age and sex, but they are mostly BMI mediated. Of those, three CpG sites flanking the leading CpG site annotated to *AKT2* and one flanking the CpG site annotated to *C14orf182* were significantly associated with incident T2D. One CpG site for *AKT2* and one for *C14orf182* were still significant after adjustment for BMI. Furthermore, the methylation levels of in total 15 CpG sites were genome-wide significantly associated with fasting glucose, 2-hour glucose, fasting insulin, and HOMA-IR, independent of age, sex, and estimated white blood cell proportions, but were widely BMI mediated. For each of these glycemic parameters, association with cg06500161 annotated to *ABCG1* can be observed. Findings for methylation patterns at the *ABCG1* gene are consistent with previous evidence that this gene is involved in HDL/LDL metabolism and diabetes in

a direct or indirect way. In both studies an accumulation of annotated genes in pathways that can be linked to diabetes can be found.

From the findings presented in this study it can be concluded that DNA methylation markers are associated with incident T2D and measures of glucose metabolism which can be measured in whole blood. These results can help to better understand the pathogenesis of T2D and related traits and can be therefore used in the future to develop biomarkers, help to predict an increased risk for this disease or develop new strategies for prevention as well as treatment of T2D.

4.6. Outlook

The findings of systematic genome-wide studies on epigenetic regulation mechanisms for incident diabetes and measures of glucose metabolism in whole blood have provided a solid basis for functional analysis to better understand how DNA methylation of the identified genes is involved in diabetes. Furthermore, the replication of these findings in insulin-sensitive tissue like adipocytes, muscle or liver, as well as diabetes-related tissues like β -cells, can be performed to investigate if whole blood reflects the epigenetic mechanism in these tissues and can therefore be used as a biomarker. Additionally, analyses between potential sequence variants and DNA methylation have to be confirmed, by for example looking for the known T2D-associated SNPs or within a defined area around significant CpG sites, to exclude that the findings are due to genetic modifications. Mendelian randomization or time-series studies should be performed to allow a conclusion of causality in the cross-sectional study. Also meta-analyses have to be performed to raise the sample size and thereby the power to detect further CpG sites which could not be detected in this study.

5. Literature

1. Internal Diabetes Federation, *IDF Diabetes Atlas*, 2013, International Diabetes Federation: Brussels,Belgium.
2. Tamayo, T., et al., *Diabetes in Europe: an update*. *Diabetes Res Clin Pract*, 2014. **103**(2): p. 206-17.
3. World Health Organisation *Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications* 1999. 1-66.
4. ADA, *Diagnosis and classification of diabetes mellitus*. *Diabetes Care*, 2011. **34 Suppl 1**: p. S62-9.
5. Ali, O., *Genetics of type 2 diabetes*. *World J Diabetes*, 2013. **4**(4): p. 114-23.
6. Nolan, C.J., P. Damm, and M. Prentki, *Type 2 diabetes across generations: from pathophysiology to prevention and management*. *Lancet*, 2011. **378**(9786): p. 169-81.
7. Kirchhoff, K., et al., *Genetik des Diabetes mellitus Typ 2*. *Med Welt*, 2008. **59**: p. 1-6.
8. Wheeler, E. and I. Barroso, *Genome-wide association studies and type 2 diabetes*. *Brief Funct Genomics*, 2011. **10**(2): p. 52-60.
9. Duncan, B.B., et al., *Low-grade systemic inflammation and the development of type 2 diabetes: the atherosclerosis risk in communities study*. *Diabetes*, 2003. **52**(7): p. 1799-805.
10. Franks, P.W., *Gene x environment interactions in type 2 diabetes*. *Curr Diab Rep*, 2011. **11**(6): p. 552-61.
11. Kriebel, J., H. Grallert, and T. Illig, *Typ-2-diabetes-assoziierte Gene*. *Der Diabetologe*, 2012. **8**: p. 26-34. "With kind permission of Springer Science+Business Media"
12. ADA, *2. Classification and Diagnosis of Diabetes*. *Diabetes Care*, 2015. **38**(Supplement 1): p. S8-S16.
13. ADA, *Standards of medical care in diabetes--2014*. *Diabetes Care*, 2014. **37 Suppl 1**: p. S14-80.
14. Diabetes Prevention Program Research, G., *HbA1c as a Predictor of Diabetes and as an Outcome in the Diabetes Prevention Program: A Randomized Clinical Trial*. *Diabetes Care*, 2015. **38**(1): p. 51-8.
15. World Health Organisation *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus* 2011. 1-25.
16. Picon, M.J., et al., *Hemoglobin A1c versus oral glucose tolerance test in postpartum diabetes screening*. *Diabetes Care*, 2012. **35**(8): p. 1648-53.
17. Kowall, B., W. Rathmann, and R. Landgraf, *Is HbA1c a valid and feasible tool for the diagnosis of diabetes?* *Diabetes Res Clin Pract*, 2011. **93**(3): p. 314-6.
18. Herder, C., *Genetische Studien zum Typ-2-Diabetes*. *Der Diabetologe*, 2010. **6**(3): p. 203-209.
19. Park, K.S., *The search for genetic risk factors of type 2 diabetes mellitus*. *Diabetes Metab J*, 2011. **35**(1): p. 12-22.
20. Herder, C., M. Karakas, and W. Koenig, *Biomarkers for the prediction of type 2 diabetes and cardiovascular disease*. *Clin Pharmacol Ther*, 2011. **90**(1): p. 52-66.
21. Drong, A.W., C.M. Lindgren, and M.I. McCarthy, *The genetic and epigenetic basis of type 2 diabetes and obesity*. *Clin Pharmacol Ther*, 2012. **92**(6): p. 707-15.
22. Qin, J., et al., *A metagenome-wide association study of gut microbiota in type 2 diabetes*. *Nature*, 2012. **490**(7418): p. 55-60.
23. Glans, F., et al., *Immigrants from the Middle-East have a different form of Type 2 diabetes compared with Swedish patients*. *Diabet Med*, 2008. **25**(3): p. 303-7.
24. Ayas, N.T., et al., *A prospective study of self-reported sleep duration and incident diabetes in women*. *Diabetes Care*, 2003. **26**(2): p. 380-4.
25. Basile, K.J., et al., *Genetic susceptibility to type 2 diabetes and obesity: follow-up of findings from genome-wide association studies*. *Int J Endocrinol*, 2014. **2014**: p. 769671.
26. Groop, L. and F. Pociot, *Genetics of diabetes--are we missing the genes or the disease?* *Mol Cell Endocrinol*, 2014. **382**(1): p. 726-39.

27. Grant, S.F., et al., *Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes*. Nat Genet, 2006. **38**(3): p. 320-3.
28. Unoki, H., et al., *SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations*. Nat Genet, 2008. **40**(9): p. 1098-102.
29. Yasuda, K., et al., *Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus*. Nat Genet, 2008. **40**(9): p. 1092-7.
30. McCarthy, M.I., *Genomics, type 2 diabetes, and obesity*. N Engl J Med, 2010. **363**(24): p. 2339-50.
31. Yamauchi, T., et al., *A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B*. Nat Genet, 2010. **42**(10): p. 864-8.
32. Imamura, M. and S. Maeda, *Genetics of type 2 diabetes: the GWAS era and future perspectives [Review]*. Endocr J, 2011. **58**(9): p. 723-39.
33. Travers, M.E. and M.I. McCarthy, *Type 2 diabetes and obesity: genomics and the clinic*. Hum Genet, 2011. **130**(1): p. 41-58.
34. Cauchi, S., et al., *TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis*. J Mol Med (Berl), 2007. **85**(7): p. 777-82.
35. McCarthy, M.I., *The importance of global studies of the genetics of type 2 diabetes*. Diabetes Metab J, 2011. **35**(2): p. 91-100.
36. DIAGRAM, et al., *Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility*. Nat Genet, 2014. **46**(3): p. 234-44.
37. Sun, X., W. Yu, and C. Hu, *Genetics of type 2 diabetes: insights into the pathogenesis and its clinical application*. Biomed Res Int, 2014. **2014**: p. 926713.
38. Marullo, L., J.S. El-Sayed Moustafa, and I. Prokopenko, *Insights into the genetic susceptibility to type 2 diabetes from genome-wide association studies of glycaemic traits*. Curr Diab Rep, 2014. **14**(11): p. 551.
39. Voight, B.F., et al., *Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis*. Nat Genet, 2010. **42**(7): p. 579-89.
40. Lyssenko, V., et al., *Clinical risk factors, DNA variants, and the development of type 2 diabetes*. N Engl J Med, 2008. **359**(21): p. 2220-32.
41. Poulsen, P., et al., *Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study*. Diabetologia, 1999. **42**(2): p. 139-45.
42. Poulsen, P., et al., *Heritability of insulin secretion, peripheral and hepatic insulin action, and intracellular glucose partitioning in young and old Danish twins*. Diabetes, 2005. **54**(1): p. 275-83.
43. Schousboe, K., et al., *Twin study of genetic and environmental influences on glucose tolerance and indices of insulin sensitivity and secretion*. Diabetologia, 2003. **46**(9): p. 1276-83.
44. Morris, A., et al., *Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes*. Nat Genet, 2012. **accepted**.
45. Saxena, R., et al., *Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci*. Am J Hum Genet, 2012. **90**(3): p. 410-25.
46. Gloyn, A.L. and M.I. McCarthy, *Variation across the allele frequency spectrum*. Nat Genet, 2010. **42**(8): p. 648-50.
47. Todd, J.L., et al., *The state of genome-wide association studies in pulmonary disease: a new perspective*. Am J Respir Crit Care Med, 2011. **184**(8): p. 873-80.
48. Albrechtsen, A., et al., *Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes*. Diabetologia, 2013. **56**(2): p. 298-310.
49. Zuk, O., et al., *The mystery of missing heritability: Genetic interactions create phantom heritability*. Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1193-8.
50. Marchini, J., P. Donnelly, and L.R. Cardon, *Genome-wide strategies for detecting multiple loci that influence complex diseases*. Nat Genet, 2005. **37**(4): p. 413-7.
51. Phillips, P.C., *Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems*. Nat Rev Genet, 2008. **9**(11): p. 855-67.

52. Musani, S.K., et al., *Detection of gene x gene interactions in genome-wide association studies of human population data*. Hum Hered, 2007. **63**(2): p. 67-84.
53. Kong, A., et al., *Parental origin of sequence variants associated with complex diseases*. Nature, 2009. **462**(7275): p. 868-74.
54. Small, K.S., et al., *Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes*. Nat Genet, 2011. **43**(6): p. 561-4.
55. Bergholdt, R., et al., *Identification of novel type 1 diabetes candidate genes by integrating genome-wide association data, protein-protein interactions, and human pancreatic islet gene expression*. Diabetes, 2012. **61**(4): p. 954-62.
56. Suhre, K., et al., *Human metabolic individuality in biomedical and pharmaceutical research*. Nature, 2011. **477**(7362): p. 54-60.
57. Brorsson, C., et al., *The type 1 diabetes - HLA susceptibility interactome--identification of HLA genotype-specific disease genes for type 1 diabetes*. PLoS One, 2010. **5**(3): p. e9576.
58. Kriebel, J., T. Illig, and H. Grallert, *Epigenetische Prozesse beim Typ-2-Diabetes. Beitrag zum Verständnis der genetischen Prädisposition*. Diabetologe, 2013. **9**(3): p. 243-250. "With kind permission of Springer Science+Business Media"
59. Holliday, R., *Epigenetics: a historical overview*. Epigenetics, 2006. **1**(2): p. 76-80.
60. Bird, A., *DNA methylation patterns and epigenetic memory*. Genes Dev, 2002. **16**(1): p. 6-21.
61. Skinner, M.K., *Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability*. Epigenetics, 2011. **6**(7): p. 838-42.
62. Langevin, S.M. and K.T. Kelsey, *The fate is not always written in the genes: epigenomics in epidemiologic studies*. Environ Mol Mutagen, 2013. **54**(7): p. 533-41.
63. Waki, H., T. Yamauchi, and T. Kadowaki, *The Epigenome and Its Role in Diabetes*. Curr Diab Rep, 2012.
64. Heard, E. and R.A. Martienssen, *Transgenerational epigenetic inheritance: myths and mechanisms*. Cell, 2014. **157**(1): p. 95-109.
65. Cubas, P., C. Vincent, and E. Coen, *An epigenetic mutation responsible for natural variation in floral symmetry*. Nature, 1999. **401**(6749): p. 157-61.
66. Dunn, G.A. and T.L. Bale, *Maternal high-fat diet effects on third-generation female body size via the paternal lineage*. Endocrinology, 2011. **152**(6): p. 2228-36.
67. Morgan, H.D., et al., *Epigenetic inheritance at the agouti locus in the mouse*. Nat Genet, 1999. **23**(3): p. 314-8.
68. Heijmans, B.T., et al., *Persistent epigenetic differences associated with prenatal exposure to famine in humans*. Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17046-9.
69. Painter, R.C., et al., *Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life*. BJOG, 2008. **115**(10): p. 1243-9.
70. Veenendaal, M.V., et al., *Transgenerational effects of prenatal exposure to the 1944-45 Dutch famine*. BJOG, 2013. **120**(5): p. 548-53.
71. Hou, L., et al., *Environmental chemical exposures and human epigenetics*. Int J Epidemiol, 2012. **41**(1): p. 79-105.
72. Kaelin, W.G., Jr. and S.L. McKnight, *Influence of metabolism on epigenetics and disease*. Cell, 2013. **153**(1): p. 56-69.
73. Berger, S.L., *The complex language of chromatin regulation during transcription*. Nature, 2007. **447**(7143): p. 407-12.
74. Griffith, J.S. and H.R. Mahler, *DNA ticketing theory of memory*. Nature, 1969. **223**(5206): p. 580-2.
75. Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development*. Science, 1975. **187**(4173): p. 226-32.
76. Riggs, A.D., *X inactivation, differentiation, and DNA methylation*. Cytogenet Cell Genet, 1975. **14**(1): p. 9-25.
77. Ulrey, C.L., et al., *The impact of metabolism on DNA methylation*. Hum Mol Genet, 2005. **14 Spec No 1**: p. R139-47.
78. Grewal, S.I. and D. Moazed, *Heterochromatin and epigenetic control of gene expression*. Science, 2003. **301**(5634): p. 798-802.

79. Reik, W., W. Dean, and J. Walter, *Epigenetic reprogramming in mammalian development*. Science, 2001. **293**(5532): p. 1089-93.
80. Lippman, Z., et al., *Role of transposable elements in heterochromatin and epigenetic control*. Nature, 2004. **430**(6998): p. 471-6.
81. Reik, W., *Stability and flexibility of epigenetic gene regulation in mammalian development*. Nature, 2007. **447**(7143): p. 425-32.
82. Straussman, R., et al., *Developmental programming of CpG island methylation profiles in the human genome*. Nat Struct Mol Biol, 2009. **16**(5): p. 564-71.
83. Sandoval, J., et al., *Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome*. Epigenetics, 2011. **6**(6): p. 692-702.
84. Bibikova, M., et al., *High density DNA methylation array with single CpG site resolution*. Genomics, 2011. **98**(4): p. 288-95.
85. Baccarelli, A. and V. Bollati, *Epigenetics and environmental chemicals*. Curr Opin Pediatr, 2009. **21**(2): p. 243-51.
86. Choudhuri, S., Y. Cui, and C.D. Klaassen, *Molecular targets of epigenetic regulation and effectors of environmental influences*. Toxicol Appl Pharmacol, 2010. **245**(3): p. 378-93.
87. Irizarry, R.A., et al., *The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores*. Nat Genet, 2009. **41**(2): p. 178-86.
88. Doi, A., et al., *Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts*. Nat Genet, 2009. **41**(12): p. 1350-3.
89. Baylin, S.B. and P.A. Jones, *A decade of exploring the cancer epigenome - biological and translational implications*. Nat Rev Cancer, 2011. **11**(10): p. 726-34.
90. Rivera, R.M. and L.B. Bennett, *Epigenetics in humans: an overview*. Curr Opin Endocrinol Diabetes Obes, 2010. **17**(6): p. 493-9.
91. Ramchandani, S., et al., *DNA methylation is a reversible biological signal*. Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6107-12.
92. Handy, D.E., R. Castro, and J. Loscalzo, *Epigenetic modifications: basic mechanisms and role in cardiovascular disease*. Circulation, 2011. **123**(19): p. 2145-56.
93. Condliffe, D., et al., *Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain*. Neurobiol Aging, 2014. **35**(8): p. 1850-4.
94. Ito, S., et al., *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification*. Nature, 2010. **466**(7310): p. 1129-33.
95. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*. Science, 2009. **324**(5929): p. 930-5.
96. Branco, M.R., G. Ficz, and W. Reik, *Uncovering the role of 5-hydroxymethylcytosine in the epigenome*. Nat Rev Genet, 2012. **13**(1): p. 7-13.
97. Wang, F., et al., *Genome-wide loss of 5-hmC is a novel epigenetic feature of Huntington's disease*. Hum Mol Genet, 2013. **22**(18): p. 3641-53.
98. Shen, J.C., W.M. Rideout, 3rd, and P.A. Jones, *The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA*. Nucleic Acids Res, 1994. **22**(6): p. 972-6.
99. Cooper, D.N., et al., *Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides*. Hum Genomics, 2010. **4**(6): p. 406-10.
100. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences*. Nature, 2009. **462**(7271): p. 315-22.
101. Rakyan, V.K., et al., *Epigenome-wide association studies for common human diseases*. Nat Rev Genet, 2011. **12**(8): p. 529-41.
102. Zeilinger, S., et al., *Tobacco smoking leads to extensive genome-wide changes in DNA methylation*. PLoS One, 2013. **8**(5): p. e63812.
103. Cao-Lei, L., et al., *DNA methylation signatures triggered by prenatal maternal stress exposure to a natural disaster: Project Ice Storm*. PLoS One, 2014. **9**(9): p. e107653.
104. Sofer, T., et al., *Exposure to airborne particulate matter is associated with methylation pattern in the asthma pathway*. Epigenomics, 2013. **5**(2): p. 147-54.

105. Seki, Y., et al., *Minireview: Epigenetic programming of diabetes and obesity: animal models*. *Endocrinology*, 2012. **153**(3): p. 1031-8.
106. Slomko, H., H.J. Heo, and F.H. Einstein, *Minireview: Epigenetics of obesity and diabetes in humans*. *Endocrinology*, 2012. **153**(3): p. 1025-30.
107. Yokomori, N., M. Tawata, and T. Onaya, *DNA demethylation during the differentiation of 3T3-L1 cells affects the expression of the mouse GLUT4 gene*. *Diabetes*, 1999. **48**(4): p. 685-90.
108. Fujiki, K., et al., *Expression of the peroxisome proliferator activated receptor gamma gene is repressed by DNA methylation in visceral adipose tissue of mouse models of diabetes*. *BMC Biol*, 2009. **7**: p. 38.
109. Kuroda, A., et al., *Insulin gene expression is regulated by DNA methylation*. *PLoS One*, 2009. **4**(9): p. e6953.
110. Barres, R., et al., *Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density*. *Cell Metab*, 2009. **10**(3): p. 189-98.
111. Ling, C., et al., *Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion*. *Diabetologia*, 2008. **51**(4): p. 615-22.
112. Ribel-Madsen, R., et al., *Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes*. *PLoS One*, 2012. **7**(12): p. e51302.
113. Bell, C.G., et al., *Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus*. *PLoS One*, 2010. **5**(11): p. e14040.
114. Toperoff, G., et al., *Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood*. *Hum Mol Genet*, 2012. **21**(2): p. 371-83.
115. Yang, B.T., et al., *Increased DNA Methylation and Decreased Expression of PDX-1 in Pancreatic Islets from Patients with Type 2 Diabetes*. *Mol Endocrinol*, 2012. **26**(7): p. 1203-12.
116. Dayeh, T., 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**(3): p. e1004160.
117. Hall, E., et al., *Effects of palmitate on genome-wide mRNA expression and DNA methylation patterns in human pancreatic islets*. *BMC Med*, 2014. **12**: p. 103.
118. Dayeh, T.A., et al., *Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets*. *Diabetologia*, 2013. **56**(5): p. 1036-46.
119. Volkmar, M., et al., *DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients*. *EMBO J*, 2012. **31**(6): p. 1405-26.
120. Nitert, M.D., 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**(12): p. 3322-32.
121. Holle, R., et al., *KORA--a research platform for population based health research*. *Gesundheitswesen*, 2005. **67 Suppl 1**: p. S19-25.
122. Steffens, M., et al., *SNP-based analysis of genetic substructure in the German population*. *Hum Hered*, 2006. **62**(1): p. 20-9.
123. Lubin, J.H. and M.H. Gail, *Biased selection of controls for case-control analyses of cohort studies*. *Biometrics*, 1984. **40**(1): p. 63-75.
124. Robins, J.M., M.H. Gail, and J.H. Lubin, *More on "Biased selection of controls for case-control analyses of cohort studies"*. *Biometrics*, 1986. **42**(2): p. 293-9.
125. Rathmann, W., et al., *High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000*. *Diabetologia*, 2003. **46**(2): p. 182-9.
126. Rathmann, W., et al., *Hemoglobin A1c and glucose criteria identify different subjects as having type 2 diabetes in middle-aged and older populations: the KORA S4/F4 Study*. *Ann Med*, 2012. **44**(2): p. 170-7.

127. Rathmann, W., et al., *Incidence of Type 2 diabetes in the elderly German population and the effect of clinical and lifestyle risk factors: KORA S4/F4 cohort study*. *Diabet Med*, 2009. **26**(12): p. 1212-9.
128. Eng, B., P. Ainsworth, and J.S. Wayne, *Anomalous migration of PCR products using nondenaturing polyacrylamide gel electrophoresis: the amelogenin sex-typing system*. *J Forensic Sci*, 1994. **39**(6): p. 1356-9.
129. Hayatsu, H., *Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis--a personal account*. *Proc Jpn Acad Ser B Phys Biol Sci*, 2008. **84**(8): p. 321-30.
130. Laird, P.W., *Principles and challenges of genomewide DNA methylation analysis*. *Nat Rev Genet*, 2010. **11**(3): p. 191-203.
131. Dedeurwaerder, S., et al., *Evaluation of the Infinium Methylation 450K technology*. *Epigenomics*, 2011. **3**(6): p. 771-84.
132. Ehrich, M., et al., *Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry*. *Proc Natl Acad Sci U S A*, 2005. **102**(44): p. 15785-90.
133. Hillenkamp, F., et al., *Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers*. *Anal Chem*, 1991. **63**(24): p. 1193A-1203A.
134. Touleimat, N. and J. Tost, *Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation*. *Epigenomics*, 2012. **4**(3): p. 325-41.
135. Teschendorff, A.E., et al., *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data*. *Bioinformatics*, 2013. **29**(2): p. 189-96.
136. Team, R.C., *R: A Language and Environment for Statistical Computing*. <http://www.r-project.org/>, 2012.
137. Chen, Y.A., et al., *Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray*. *Epigenetics*, 2013. **8**(2): p. 203-9.
138. Price, M.E., et al., *Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array*. *Epigenetics Chromatin*, 2013. **6**(1): p. 4.
139. Marabita, F., et al., *An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform*. *Epigenetics*, 2013. **8**(3): p. 333-46.
140. Therneau, T.M., *A Package for Survival Analysis in S*. <http://CRAN.R-project.org/package=survival>, 2013. **R package version 2.37.4**.
141. Pinheiro, J., et al., *Linear and Nonlinear Mixed Effects Models*.
142. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution*. *BMC Bioinformatics*, 2012. **13**: p. 86.
143. Schurmann, C., et al., *Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium*. *PLoS One*, 2012. **7**(12): p. e50938.
144. Guo, T., et al., *Polymorphism of rs873308 near the transmembrane protein 57 gene is associated with serum lipid levels*. *Biosci Rep*, 2014.
145. Aulchenko, Y.S., et al., *Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts*. *Nat Genet*, 2009. **41**(1): p. 47-55.
146. Jung, H. and I. Choi, *Thioredoxin-interacting protein, hematopoietic stem cells, and hematopoiesis*. *Curr Opin Hematol*, 2014. **21**(4): p. 265-70.
147. Roberts, S., et al., *Functional involvement of PHOSPHO1 in matrix vesicle-mediated skeletal mineralization*. *J Bone Miner Res*, 2007. **22**(4): p. 617-27.
148. Jiang, Z.Y., et al., *Insulin signaling through Akt/protein kinase B analyzed by small interfering RNA-mediated gene silencing*. *Proc Natl Acad Sci U S A*, 2003. **100**(13): p. 7569-74.
149. Takenaka, N., et al., *Role of the guanine nucleotide exchange factor in Akt2-mediated plasma membrane translocation of GLUT4 in insulin-stimulated skeletal muscle*. *Cell Signal*, 2014. **26**(11): p. 2460-9.

150. Ng, Y., et al., *Rapid activation of Akt2 is sufficient to stimulate GLUT4 translocation in 3T3-L1 adipocytes*. Cell Metab, 2008. **7**(4): p. 348-56.
151. McCurdy, C.E. and G.D. Cartee, *Akt2 is essential for the full effect of calorie restriction on insulin-stimulated glucose uptake in skeletal muscle*. Diabetes, 2005. **54**(5): p. 1349-56.
152. Schultze, S.M., et al., *Reduced hepatic lipid content in Pten-haplodeficient mice because of enhanced AKT2/PKBbeta activation in skeletal muscle*. Liver Int, 2014.
153. Chen, W.S., et al., *Leptin deficiency and beta-cell dysfunction underlie type 2 diabetes in compound Akt knockout mice*. Mol Cell Biol, 2009. **29**(11): p. 3151-62.
154. Leavens, K.F., et al., *Akt2 is required for hepatic lipid accumulation in models of insulin resistance*. Cell Metab, 2009. **10**(5): p. 405-18.
155. Garofalo, R.S., et al., *Severe diabetes, age-dependent loss of adipose tissue, and mild growth deficiency in mice lacking Akt2/PKB beta*. J Clin Invest, 2003. **112**(2): p. 197-208.
156. Sylow, L., et al., *Akt and Rac1 signaling are jointly required for insulin-stimulated glucose uptake in skeletal muscle and downregulated in insulin resistance*. Cell Signal, 2014. **26**(2): p. 323-31.
157. Lu, M., et al., *Insulin regulates liver metabolism in vivo in the absence of hepatic Akt and Foxo1*. Nat Med, 2012. **18**(3): p. 388-95.
158. Wan, M., et al., *A noncanonical, GSK3-independent pathway controls postprandial hepatic glycogen deposition*. Cell Metab, 2013. **18**(1): p. 99-105.
159. Sun, W., et al., *Renal improvement by zinc in diabetic mice is associated with glucose metabolism signaling mediated by metallothionein and Akt, but not Akt2*. Free Radic Biol Med, 2014. **68**: p. 22-34.
160. Cho, H., et al., *Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta)*. Science, 2001. **292**(5522): p. 1728-31.
161. Liu, Z., et al., *Hypoglycemic Activity and Antioxidative Stress of Extracts and Corymbiferin from Swertia bimaculata In Vitro and In Vivo*. Evid Based Complement Alternat Med, 2013. **2013**: p. 125416.
162. Otieno, C.J., et al., *Mapping and association studies of diabetes related genes in the pig*. Anim Genet, 2005. **36**(1): p. 36-42.
163. Fischer-Posovszky, P., et al., *Differential function of Akt1 and Akt2 in human adipocytes*. Mol Cell Endocrinol, 2012. **358**(1): p. 135-43.
164. Pettersson, A.M., et al., *LXR is a negative regulator of glucose uptake in human adipocytes*. Diabetologia, 2013. **56**(9): p. 2044-54.
165. Tan, K., et al., *Analysis of genetic variation in Akt2/PKB-beta in severe insulin resistance, lipodystrophy, type 2 diabetes, and related metabolic phenotypes*. Diabetes, 2007. **56**(3): p. 714-9.
166. Albers, P.H., et al., *Human muscle fiber type-specific insulin signaling: impact of obesity and type 2 diabetes*. Diabetes, 2015. **64**(2): p. 485-97.
167. Vind, B.F., et al., *Hyperglycaemia normalises insulin action on glucose metabolism but not the impaired activation of AKT and glycogen synthase in the skeletal muscle of patients with type 2 diabetes*. Diabetologia, 2012. **55**(5): p. 1435-45.
168. Brozinick, J.T., Jr., B.R. Roberts, and G.L. Dohm, *Defective signaling through Akt-2 and -3 but not Akt-1 in insulin-resistant human skeletal muscle: potential role in insulin resistance*. Diabetes, 2003. **52**(4): p. 935-41.
169. Szendroedi, J., et al., *Role of diacylglycerol activation of PKCtheta in lipid-induced muscle insulin resistance in humans*. Proc Natl Acad Sci U S A, 2014. **111**(26): p. 9597-602.
170. Friedrichsen, M., et al., *Differential aetiology and impact of phosphoinositide 3-kinase (PI3K) and Akt signalling in skeletal muscle on in vivo insulin action*. Diabetologia, 2010. **53**(9): p. 1998-2007.
171. MacLaren, R., et al., *Influence of obesity and insulin sensitivity on insulin signaling genes in human omental and subcutaneous adipose tissue*. J Lipid Res, 2008. **49**(2): p. 308-23.

172. George, S., et al., *A family with severe insulin resistance and diabetes due to a mutation in AKT2*. Science, 2004. **304**(5675): p. 1325-8.
173. Friedrichsen, M., et al., *Akt2 influences glycogen synthase activity in human skeletal muscle through regulation of NH(2)-terminal (sites 2 + 2a) phosphorylation*. Am J Physiol Endocrinol Metab, 2013. **304**(6): p. E631-9.
174. Sun, X.Q., et al., *Contribution of the Akt2 gene to type 2 diabetes in the Chinese Han population*. Chin Med J (Engl), 2011. **124**(5): p. 725-8.
175. Han, D., et al., *Detection of differential proteomes associated with the development of type 2 diabetes in the Zucker rat model using the iTRAQ technique*. J Proteome Res, 2011. **10**(2): p. 564-77.
176. Nori, S.L., et al., *Increased nerve growth factor signaling in sensory neurons of early diabetic rats is corrected by electroacupuncture*. Evid Based Complement Alternat Med, 2013. **2013**: p. 652735.
177. Nakagaki, O., et al., *Epalrestat improves diabetic wound healing via increased expression of nerve growth factor*. Exp Clin Endocrinol Diabetes, 2013. **121**(2): p. 84-9.
178. Aloe, L., S. Rossi, and L. Manni, *Altered expression of nerve growth factor and its receptors in the kidneys of diabetic rats*. J Nephrol, 2011. **24**(6): p. 798-805.
179. Sornelli, F., et al., *Adipose tissue-derived nerve growth factor and brain-derived neurotrophic factor: results from experimental stress and diabetes*. Gen Physiol Biophys, 2009. **28 Spec No**: p. 179-83.
180. Pierucci, D., et al., *NGF-withdrawal induces apoptosis in pancreatic beta cells in vitro*. Diabetologia, 2001. **44**(10): p. 1281-95.
181. Park, S.W., et al., *Angiopoietin 2 induces pericyte apoptosis via alpha3beta1 integrin signaling in diabetic retinopathy*. Diabetes, 2014. **63**(9): p. 3057-68.
182. Feral, C.C., et al., *Blockade of alpha4 integrin signaling ameliorates the metabolic consequences of high-fat diet-induced obesity*. Diabetes, 2008. **57**(7): p. 1842-51.
183. Narayanan, K., et al., *Extracellular matrix-mediated differentiation of human embryonic stem cells: differentiation to insulin-secreting beta cells*. Tissue Eng Part A, 2014. **20**(1-2): p. 424-33.
184. Krishnamurthy, M., et al., *Integrin {alpha}3, but not {beta}1, regulates islet cell survival and function via PI3K/Akt signaling pathways*. Endocrinology, 2011. **152**(2): p. 424-35.
185. Osorio-Fuentealba, C., et al., *Electrical stimuli release ATP to increase GLUT4 translocation and glucose uptake via PI3Kgamma-Akt-AS160 in skeletal muscle cells*. Diabetes, 2013. **62**(5): p. 1519-26.
186. Cha-Molstad, H., et al., *Glucose-stimulated expression of Txnip is mediated by carbohydrate response element-binding protein, p300, and histone H4 acetylation in pancreatic beta cells*. J Biol Chem, 2009. **284**(25): p. 16898-905.
187. Chen, J., et al., *Thioredoxin-interacting protein deficiency induces Akt/Bcl-xL signaling and pancreatic beta-cell mass and protects against diabetes*. FASEB J, 2008. **22**(10): p. 3581-94.
188. Xu, G., et al., *Thioredoxin-interacting protein regulates insulin transcription through microRNA-204*. Nat Med, 2013. **19**(9): p. 1141-6.
189. Jo, S.H., et al., *Txnip contributes to impaired glucose tolerance by upregulating the expression of genes involved in hepatic gluconeogenesis in mice*. Diabetologia, 2013. **56**(12): p. 2723-32.
190. Chai, T.F., et al., *A potential mechanism of metformin-mediated regulation of glucose homeostasis: inhibition of Thioredoxin-interacting protein (Txnip) gene expression*. Cell Signal, 2012. **24**(8): p. 1700-5.
191. Pedroso, J.A., et al., *Inactivation of SOCS3 in leptin receptor-expressing cells protects mice from diet-induced insulin resistance but does not prevent obesity*. Mol Metab, 2014. **3**(6): p. 608-18.
192. Jorgensen, S.B., et al., *Deletion of skeletal muscle SOCS3 prevents insulin resistance in obesity*. Diabetes, 2013. **62**(1): p. 56-64.
193. Shi, H., et al., *Overexpression of suppressor of cytokine signaling 3 in adipose tissue causes local but not systemic insulin resistance*. Diabetes, 2006. **55**(3): p. 699-707.

194. Mori, H., et al., *Suppression of SOCS3 expression in the pancreatic beta-cell leads to resistance to type 1 diabetes*. *Biochem Biophys Res Commun*, 2007. **359**(4): p. 952-8.
195. Harder-Lauridsen, N.M., et al., *Effect of IL-6 on the insulin sensitivity in patients with type 2 diabetes*. *Am J Physiol Endocrinol Metab*, 2014. **306**(7): p. E769-78.
196. Li, P., et al., *Genetic association analysis of 30 genes related to obesity in a European American population*. *Int J Obes (Lond)*, 2014. **38**(5): p. 724-9.
197. Fischer-Rosinsky, A., et al., *Lack of association between the tagging SNP A+930-->G of SOCS3 and type 2 diabetes mellitus: meta-analysis of four independent study populations*. *PLoS One*, 2008. **3**(12): p. e3852.
198. Zhang, J., et al., *Analysis of transcription factor Stk40 expression and function during mouse pre-implantation embryonic development*. *Mol Med Rep*, 2014. **9**(2): p. 535-40.
199. Li, L., et al., *Stk40 links the pluripotency factor Oct4 to the Erk/MAPK pathway and controls extraembryonic endoderm differentiation*. *Proc Natl Acad Sci U S A*, 2010. **107**(4): p. 1402-7.
200. Westerterp, M., et al., *ATP-binding cassette transporters, atherosclerosis, and inflammation*. *Circ Res*, 2014. **114**(1): p. 157-70.
201. Hidalgo, B., et al., *Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study*. *Diabetes*, 2014. **63**(2): p. 801-7.
202. Wilson, P.W., D.L. McGee, and W.B. Kannel, *Obesity, very low density lipoproteins, and glucose intolerance over fourteen years: The Framingham Study*. *Am J Epidemiol*, 1981. **114**(5): p. 697-704.
203. Schou, J., et al., *ABC transporter genes and risk of type 2 diabetes: a study of 40,000 individuals from the general population*. *Diabetes Care*, 2012. **35**(12): p. 2600-6.
204. von Eckardstein, A., H. Schulte, and G. Assmann, *Risk for diabetes mellitus in middle-aged Caucasian male participants of the PROCAM study: implications for the definition of impaired fasting glucose by the American Diabetes Association*. *Prospective Cardiovascular Munster*. *J Clin Endocrinol Metab*, 2000. **85**(9): p. 3101-8.
205. Kruit, J.K., et al., *Loss of both ABCA1 and ABCG1 results in increased disturbances in islet sterol homeostasis, inflammation, and impaired beta-cell function*. *Diabetes*, 2012. **61**(3): p. 659-64.
206. Sturek, J.M., et al., *An intracellular role for ABCG1-mediated cholesterol transport in the regulated secretory pathway of mouse pancreatic beta cells*. *J Clin Invest*, 2010. **120**(7): p. 2575-89.
207. Zhou, H., et al., *Associations of ATP-binding cassette transporter A1 and G1 with insulin secretion in human insulinomas*. *Pancreas*, 2012. **41**(6): p. 934-9.
208. Johansson, L.E., et al., *Differential gene expression in adipose tissue from obese human subjects during weight loss and weight maintenance*. *Am J Clin Nutr*, 2012. **96**(1): p. 196-207.
209. Mauldin, J.P., et al., *Reduced expression of ATP-binding cassette transporter G1 increases cholesterol accumulation in macrophages of patients with type 2 diabetes mellitus*. *Circulation*, 2008. **117**(21): p. 2785-92.
210. Vitto, M.F., et al., *Reversion of steatosis by SREBP-1c antisense oligonucleotide did not improve hepatic insulin action in diet-induced obesity mice*. *Horm Metab Res*, 2012. **44**(12): p. 885-90.
211. Haas, J.T., et al., *Hepatic insulin signaling is required for obesity-dependent expression of SREBP-1c mRNA but not for feeding-dependent expression*. *Cell Metab*, 2012. **15**(6): p. 873-84.
212. Na, Y.K., et al., *Effect of body mass index on global DNA methylation in healthy Korean women*. *Mol Cells*, 2014. **37**(6): p. 467-72.
213. Stepanow, S., et al., *Allele-specific, age-dependent and BMI-associated DNA methylation of human MCHR1*. *PLoS One*, 2011. **6**(5): p. e17711.
214. da Silva Aragao, R., et al., *Maternal protein restriction impairs the transcriptional metabolic flexibility of skeletal muscle in adult rat offspring*. *Br J Nutr*, 2014. **112**(3): p. 328-37.

215. Meng, Z.X., et al., *Aberrant activation of liver X receptors impairs pancreatic beta cell function through upregulation of sterol regulatory element-binding protein 1c in mouse islets and rodent cell lines*. *Diabetologia*, 2012. **55**(6): p. 1733-44.
216. Nyman, L.R., et al., *Long term effects of high fat or high carbohydrate diets on glucose tolerance in mice with heterozygous carnitine palmitoyltransferase-1a (CPT-1a) deficiency: Diet influences on CPT1a deficient mice*. *Nutr Diabetes*, 2011. **1**: p. e14.
217. Orellana-Gavalda, J.M., et al., *Molecular therapy for obesity and diabetes based on a long-term increase in hepatic fatty-acid oxidation*. *Hepatology*, 2011. **53**(3): p. 821-32.
218. Hirota, Y., et al., *Lack of association of CPT1A polymorphisms or haplotypes on hepatic lipid content or insulin resistance in Japanese individuals with type 2 diabetes mellitus*. *Metabolism*, 2007. **56**(5): p. 656-61.
219. Hotamisligil, G.S., *Inflammation and metabolic disorders*. *Nature*, 2006. **444**(7121): p. 860-7.
220. Murray, R.Z. and J.L. Stow, *Cytokine Secretion in Macrophages: SNAREs, Rabs, and Membrane Trafficking*. *Front Immunol*, 2014. **5**: p. 538.
221. Xu, H., et al., *Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance*. *J Clin Invest*, 2003. **112**(12): p. 1821-30.
222. Uysal, K.T., et al., *Protection from obesity-induced insulin resistance in mice lacking TNF-alpha function*. *Nature*, 1997. **389**(6651): p. 610-4.
223. Ramirez Alvarado, M.M. and C. Sanchez Roitz, *[Tumor necrosis factor-alpha, insulin resistance, the lipoprotein metabolism and obesity in humans]*. *Nutr Hosp*, 2012. **27**(6): p. 1751-7.
224. Wieser, V., A.R. Moschen, and H. Tilg, *Inflammation, cytokines and insulin resistance: a clinical perspective*. *Arch Immunol Ther Exp (Warsz)*, 2013. **61**(2): p. 119-25.
225. Xu, J., et al., *Maternal circulating concentrations of tumor necrosis factor-alpha, leptin, and adiponectin in gestational diabetes mellitus: a systematic review and meta-analysis*. *ScientificWorldJournal*, 2014. **2014**: p. 926932.
226. Chen, X., et al., *Type 2 diabetes mellitus control and atherosclerosis prevention in a non-obese rat model using duodenal-jejunal bypass*. *Exp Ther Med*, 2014. **8**(3): p. 856-862.
227. Volpe, C.M., et al., *The production of nitric oxide, IL-6, and TNF-alpha in palmitate-stimulated PBMCs is enhanced through hyperglycemia in diabetes*. *Oxid Med Cell Longev*, 2014. **2014**: p. 479587.
228. Alexandraki, K., et al., *Inflammatory process in type 2 diabetes: The role of cytokines*. *Ann N Y Acad Sci*, 2006. **1084**: p. 89-117.
229. Kulkarni, R.N. and C.R. Kahn, *Ephs and ephrins keep pancreatic Beta cells connected*. *Cell*, 2007. **129**(2): p. 241-3.
230. Dorrell, C., et al., *Transcriptomes of the major human pancreatic cell types*. *Diabetologia*, 2011. **54**(11): p. 2832-44.
231. Jain, R. and E. Lammert, *Cell-cell interactions in the endocrine pancreas*. *Diabetes Obes Metab*, 2009. **11 Suppl 4**: p. 159-67.
232. Konstantinova, I., et al., *EphA-Ephrin-A-mediated beta cell communication regulates insulin secretion from pancreatic islets*. *Cell*, 2007. **129**(2): p. 359-70.
233. Ramkhalawon, B., et al., *Netrin-1 promotes adipose tissue macrophage retention and insulin resistance in obesity*. *Nat Med*, 2014. **20**(4): p. 377-84.
234. Jayakumar, C., et al., *Netrin-1, a urinary proximal tubular injury marker, is elevated early in the time course of human diabetes*. *J Nephrol*, 2014. **27**(2): p. 151-7.
235. Shi, T.J., et al., *Coenzyme Q10 prevents peripheral neuropathy and attenuates neuron loss in the db/db- mouse, a type 2 diabetes model*. *Proc Natl Acad Sci U S A*, 2013. **110**(2): p. 690-5.
236. Hirata, M., et al., *Genetic defect in phospholipase Cdelta1 protects mice from obesity by regulating thermogenesis and adipogenesis*. *Diabetes*, 2011. **60**(7): p. 1926-37.
237. Fiume, R., et al., *Nuclear PLCs affect insulin secretion by targeting PPARgamma in pancreatic beta cells*. *FASEB J*, 2012. **26**(1): p. 203-10.
238. Huang, Y., et al., *The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing*. *PLoS One*, 2010. **5**(1): p. e8888.

239. Booth, M.J., et al., *Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution*. *Science*, 2012. **336**(6083): p. 934-7.
240. Stewart, S.K., et al., *oxBS-450K: A method for analysing hydroxymethylation using 450K BeadChips*. *Methods*, 2015. **72**: p. 9-15.
241. Booth, M.J., et al., *Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine*. *Nat Protoc*, 2013. **8**(10): p. 1841-51.

6. Appendix

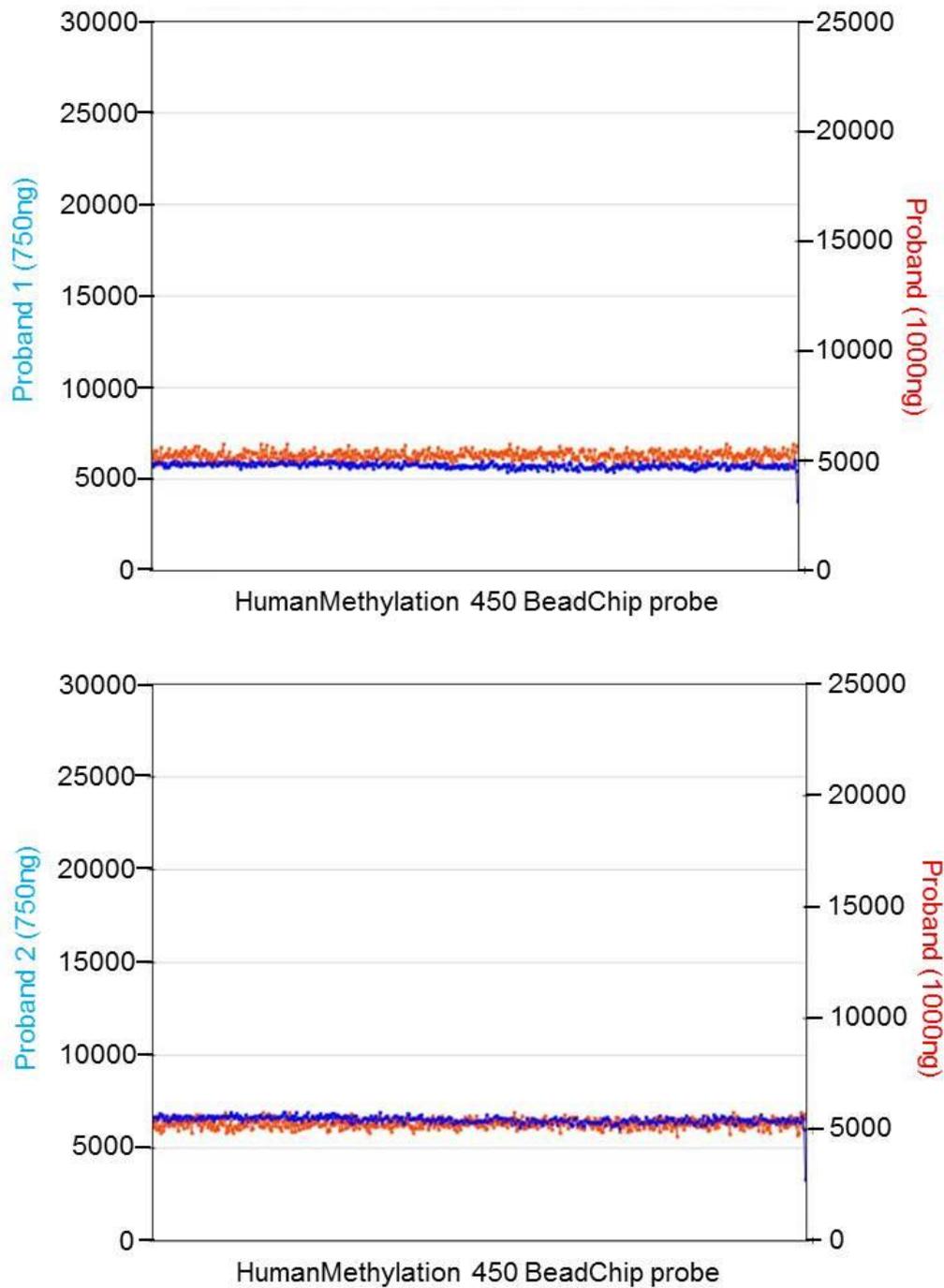
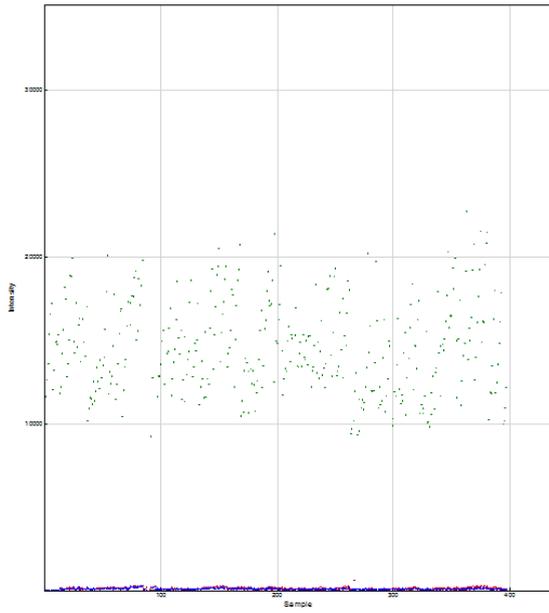
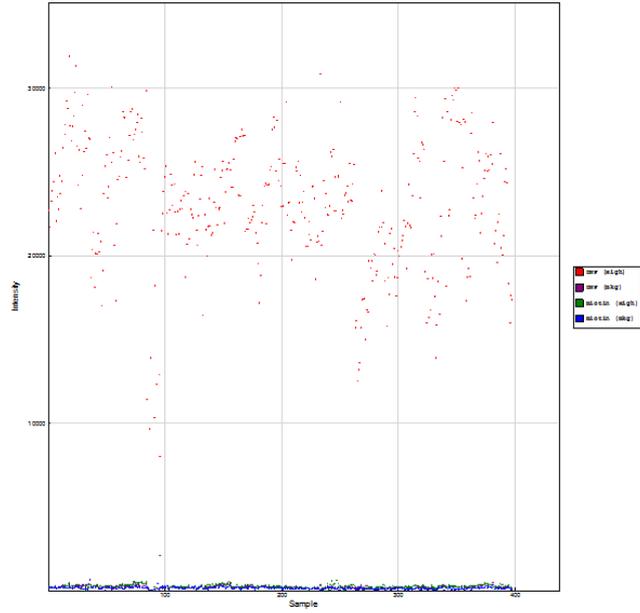


Figure 1: Line plots of results comparing 750 ng and 1,000 ng DNA. Line plots represent intensity of two different samples with 750ng and one unique sample with 1,000 ng DNA vs HumanMethylation450 BeadChip probe.

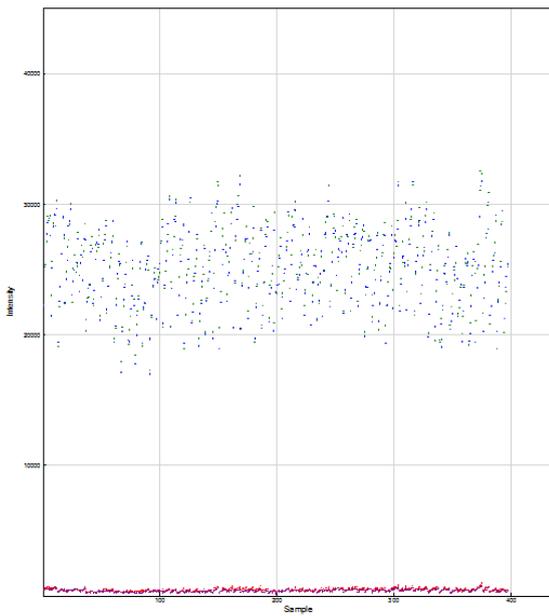
Staining Green



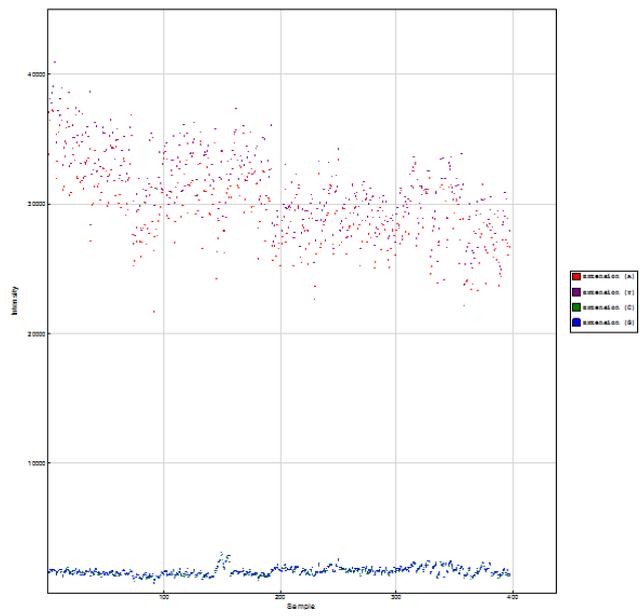
Staining Red



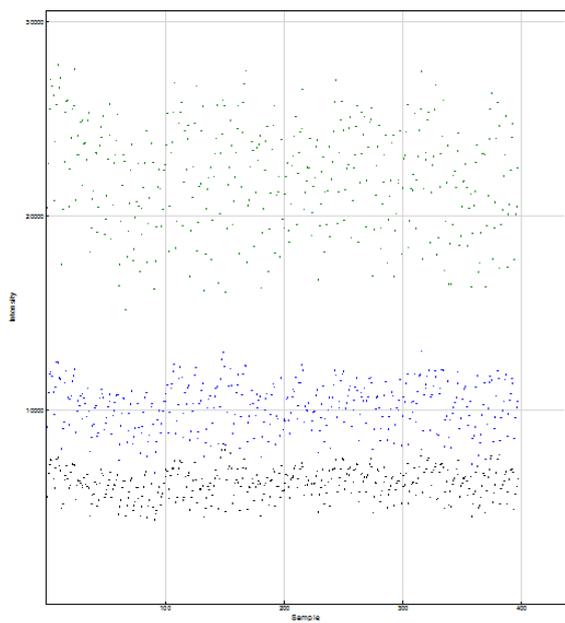
Extension Green



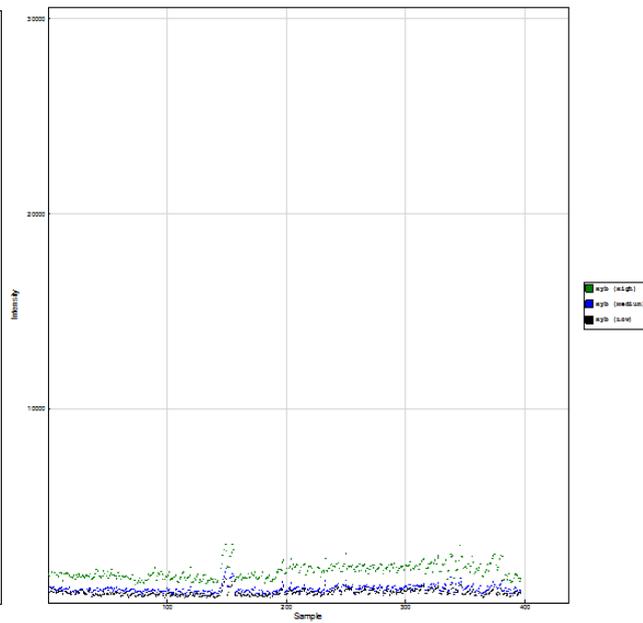
Extension Red



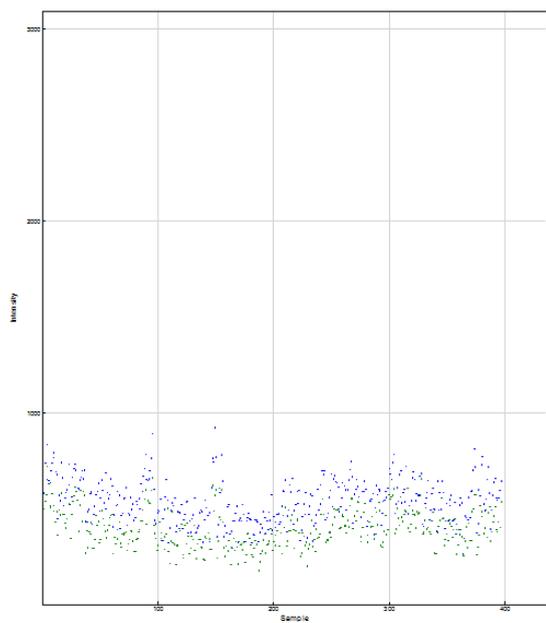
Hybridization Green



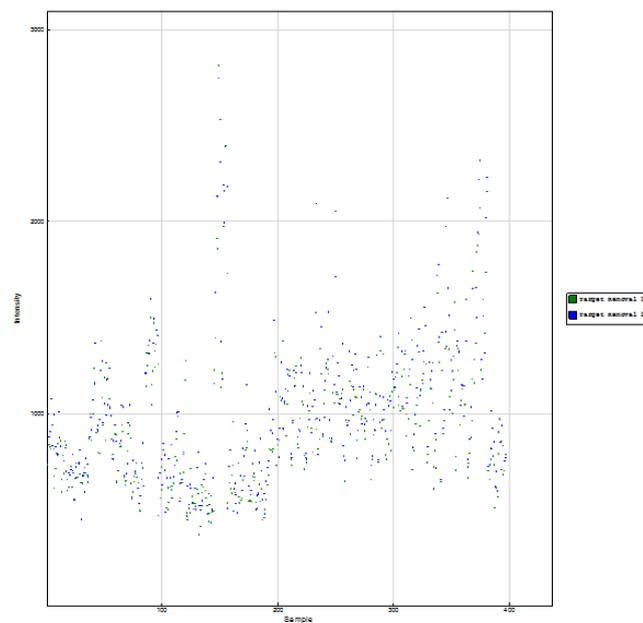
Hybridization Green



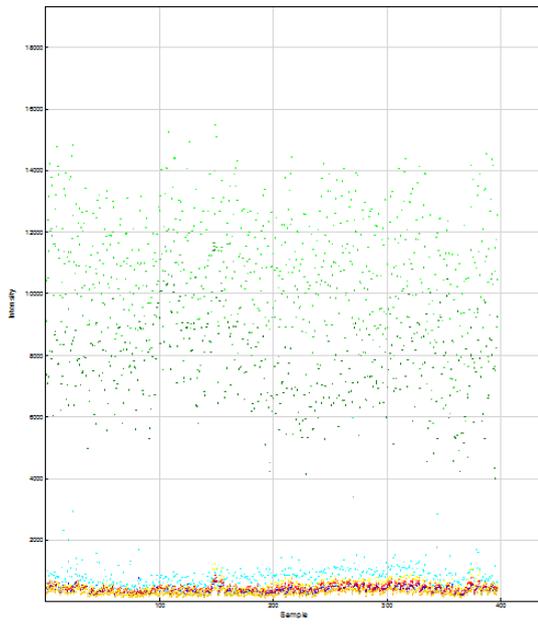
Target Removal Green



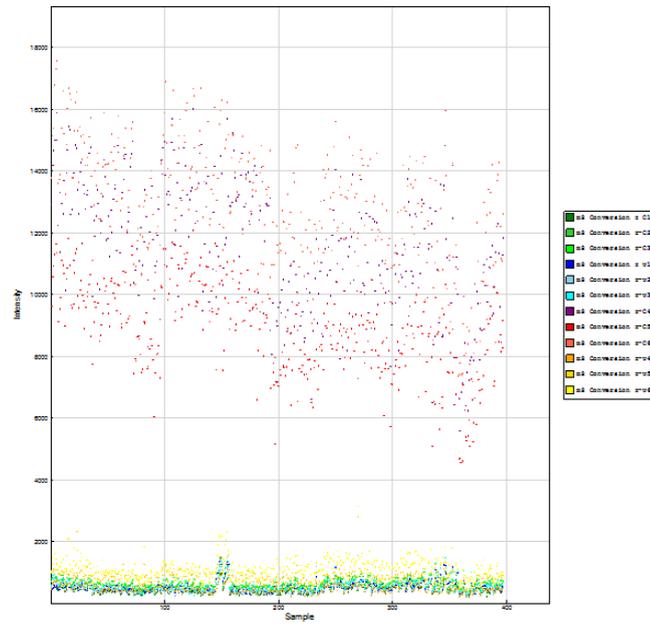
Target Removal Red



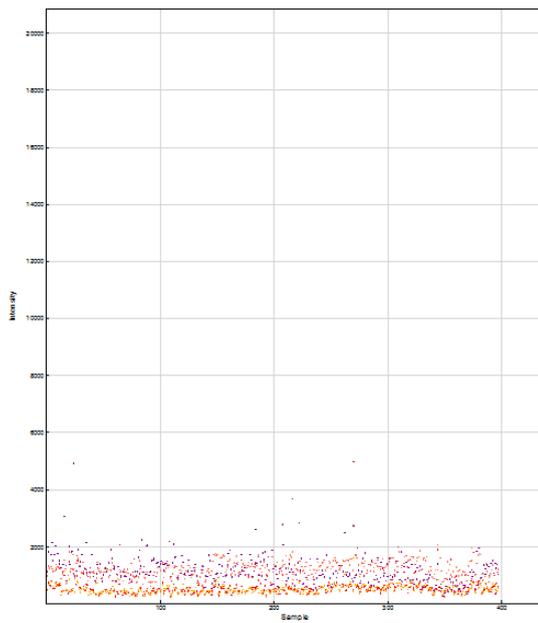
Bisulfite Conversion I Green



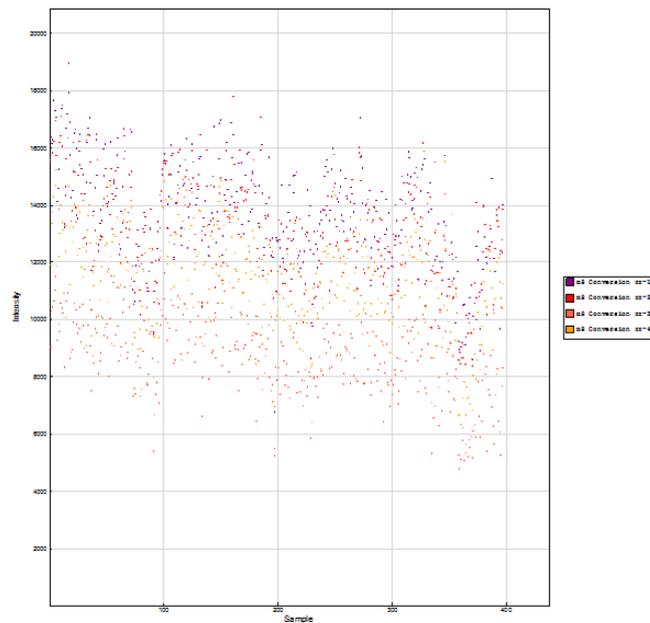
Bisulfite Conversion I Red



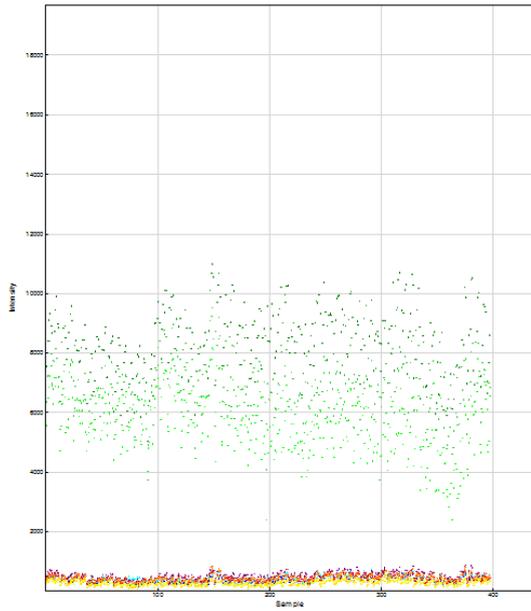
Bisulfite Conversion II Green



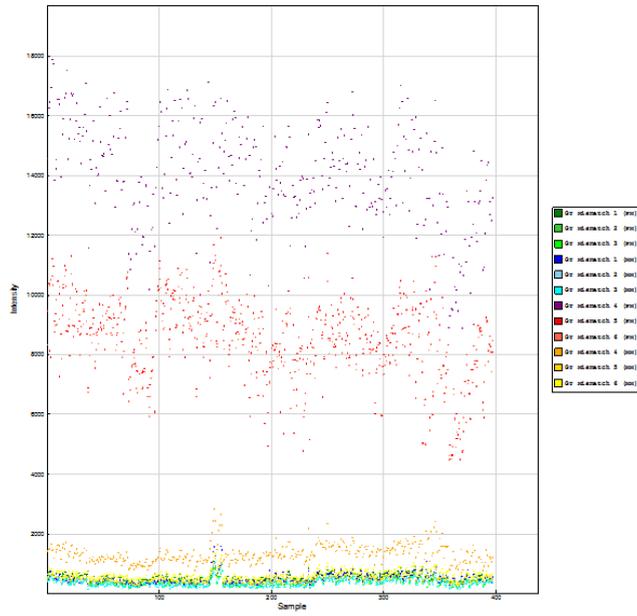
Bisulfite Conversion II Red



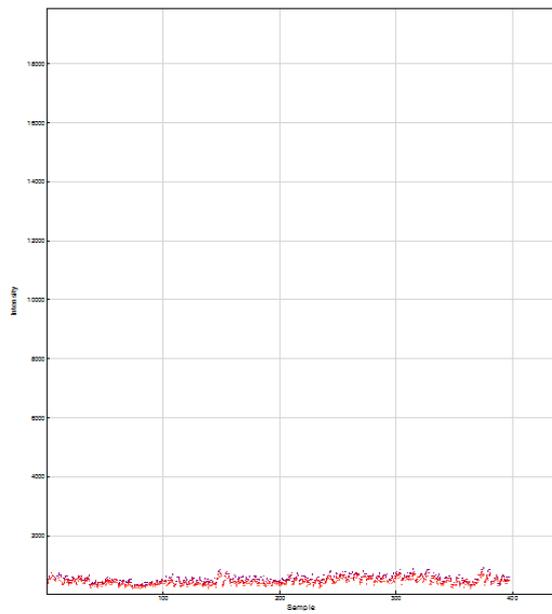
Specifity I Red



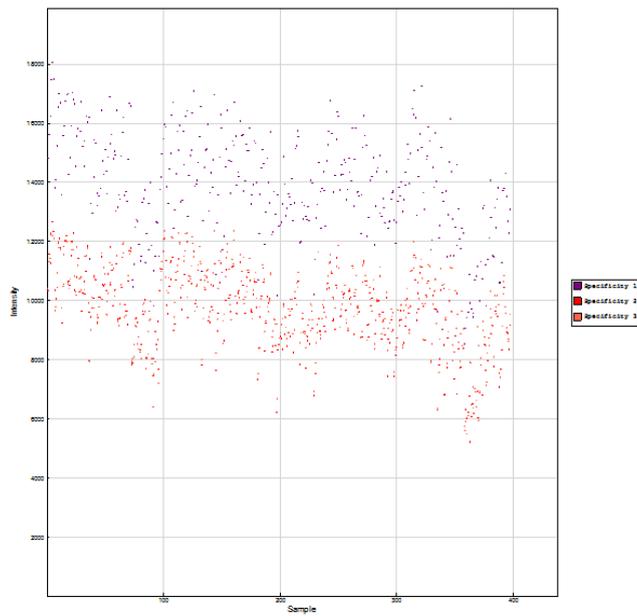
Specifity I Red



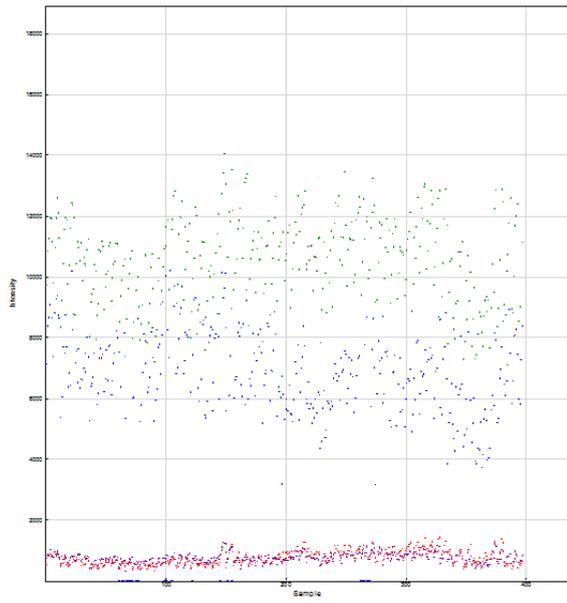
Specifity II Green



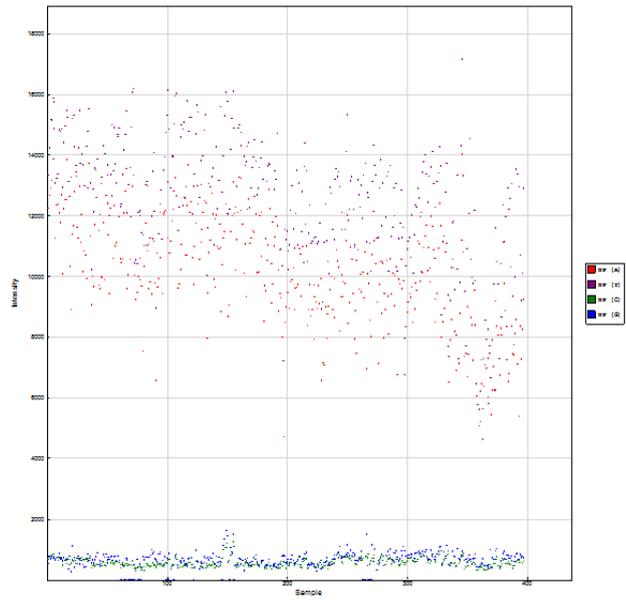
Specifity II Red



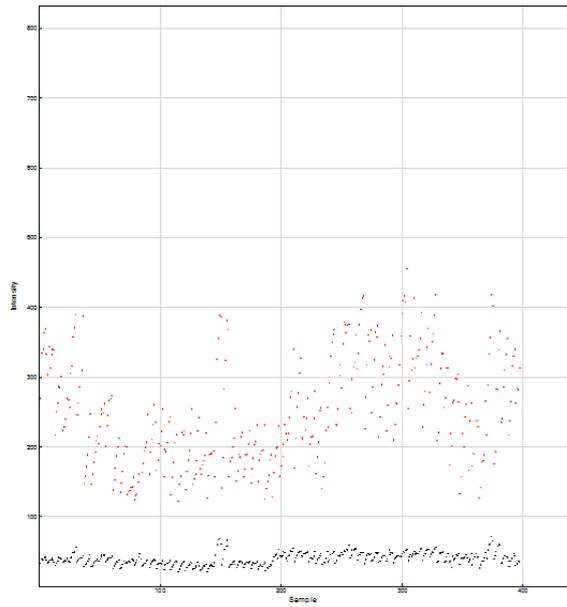
Non-Polymorphic Green



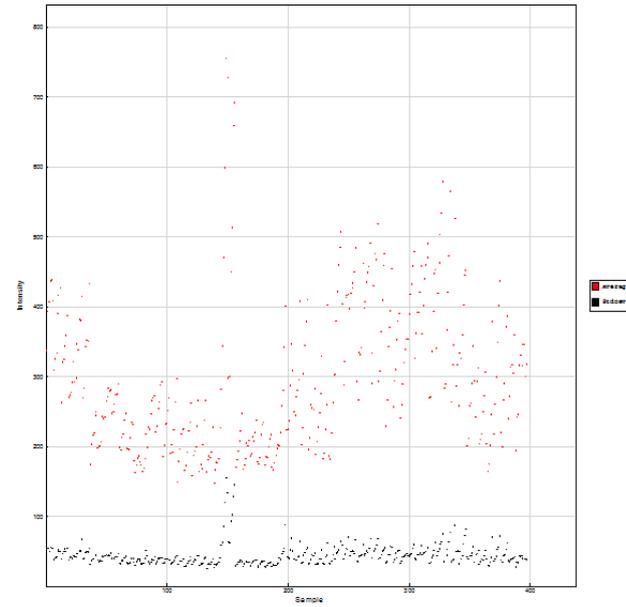
Non-Polymorphic Red



Negative Green



Negative Red



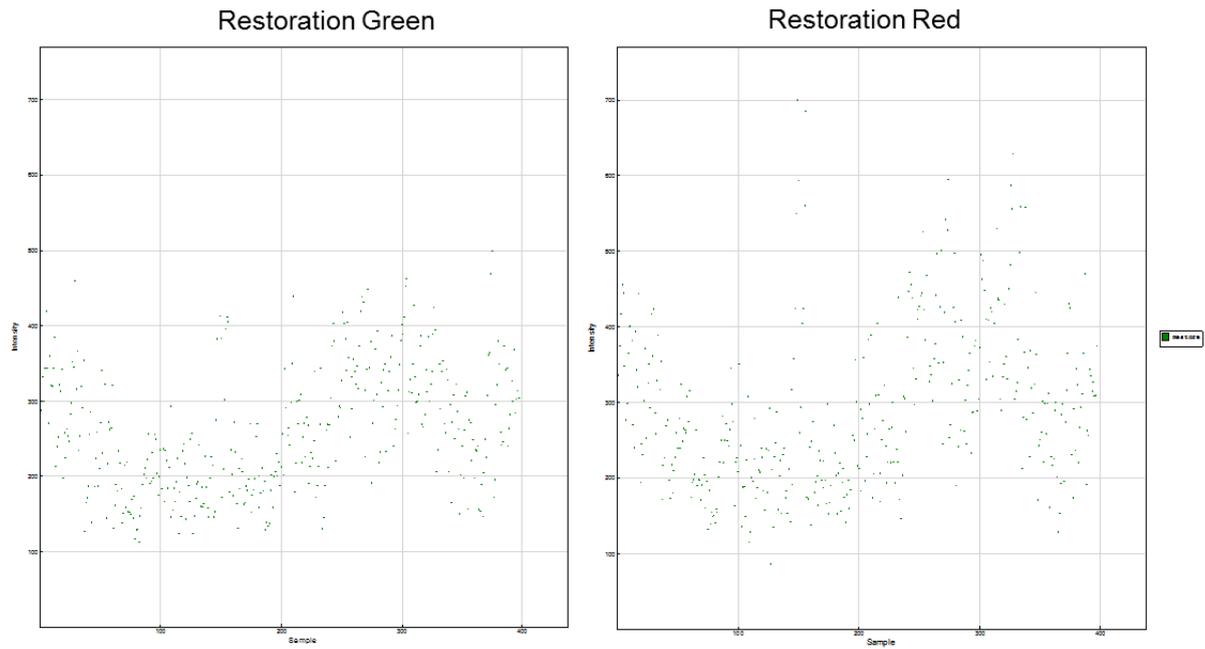


Figure 2: Screen shot from the control dashboard. x-axis represents the samples and y-axis the intensity of the samples for the corresponding control.

Table 1: Primer sequence for EpiTYPER®.

Amplicon Name	Genomic sequence for primer design	Length of target	Left primer	Right primer
A1a_cg11057824_13	TTGAACCAGTGTGGCCACCTTGCAAAGTCCCTTATGAAT GAGTTACCTACCTCACCGATACATACTCACCATATCTTGT CAAGGCTTCATATTCCTTTCCCTGTATAGTCCCCACAAAT ATACACTGTCAGCTTTCAATGCACACCAAGCGCCCATTA TTGATTCTTTGTCCCGAGTGGCTCCTCTGGAGCATATGC GTAACATCAGGGGGACATAGGGCTTTTAATTGTCAAAG TATACAAACTCAGCACGTTATTCCAAATAATCTCTGCCCC TGCTCCCTCACATTTTGATCATGCCGTATCTGTGATCTGG GGCTGGGGACTCTGGCCCTCAAGATGCCTCCAATGCTC TCGGCCCCTGCTAAGAGCCTCTGCCACACACGGGGTC ATGCGCATAACCTTAGAACTCTGGGCACGGCTAGATGAA CTGTGGCACGCACCTGATCTAGGCAGCAGCCCACTTCC CAGTTGGTCAGCCAGTGAAGGAAGGCTCAGCCCCACCA GGCCAGTGAGATTCTCCGTCTTGGCAAATTGCTTAAAGA CTTGATAGTAAAGGAGAAAGAGATAAAGCAGGCACCTGA GGCCCCGAAGCAGCAG	389	aggaagagagAGTGGGTT GTTGTTTAGATTAGGTG	cagtaatcgactcactat agggagaaggctATCA AAACTTCATATTCC TTTCCCTA
A2a_cg18514820_1	CCCGGGGTGCGGCTCCTGCAGGACTCGGTGGACTTCTC GCTGGCCGACGCCATCAACACCGAGTTCAAGAACACCC GCACCAACGAGAAAGGTGGAGCTGCAGGAGCTGAATGAC CGCTTCGCCAACTACATCGACAAGGTGCGCTTCCTGGA GCAGCAGAATAAGATCCTGCTGGCCGAGCTCGAGCAGC TCAAGGGCCAAGGCAAGTCGCGCCTGGGGGACCTCTAC GAGGAGGAGATGCGGGAGCTGCGCCGGCAGGTGGACC AGCTAACCAACGACAAAGCCCGCGTCGAGGTGGAGCGC GACAACCTGGCCGAGGACATCATGCGCCTCCGGGAGAA GTAAGGCTGCGCCCATGCAAGTAGCTGGGCCTCGGGAG	384	aggaagagagGGTGGAGT TGTAGGAGTTGAATGAT	cagtaatcgactcactat agggagaaggctCTAT CTCCCTAACAAAC CTCCC

	GGGGCTGGAGGGAGAGGGGAACGCCCCCGGCCCC GCGAGAGCTGCCACGCCCTTGGGGATGTGGCCGGGG GAGGCCTGCCAGGGAGACAGCGGAGAGCGGGGCTGTG GCTGTGGTGGCGCAGCCCCGCCAGAACCCAGACCTTG CAGTTCGCATTTCTCCTCTGTCCCCACACATTGCCCAA GGACGCTCCGTTTTCAAGTTACAGATTTCTTAAAA			
A3b_cg22876894_5	CAGGCAGTAGGTTCCGAGCCATTGCTTGTGCGTCTTCG CGGAGAGAGAGGGGGGAGATGCCCGGCGCTCCTTAC CCGGCTGGATAAGACCCTTTGGGGAACCTTGCTGCTTCG CAAGTCTTACTTGTGAGGACCCGGAGGGTTTGCCTTTGA CCCCGCCTGTGACCTTAAGAAAACCCCTTTCTACCCTGC GCCTGTCCCACAGCTGGAAGTGATAATTTAAGAATGAA TTAAGCACTCGGTGTGTGACACTGCGGCACTAAGCACTT TCTGGACTGTGTTAGGCCCATTTTACAGACGGGGAAGC TGAGGCCTCCAGAGGCGGAGTAACCTGCTCATGGTTAC AAAGCTGGTCAGCGGCGGGCAGGATTTGAATCCAGGGT TCCCGTTCCCTAATCCTCTGCTTTTACCACTTAATTTATC CTGCCTCCAAGGGCTGCTGGGAAAAGGCGGATTTCAA AGGAGAAGCAGCTAAGGAAAAGCTCTGGAAAGGGCCAA CGGAGCGGATATTCCCGGAGCCCTCTGCGAGCCACGC GCCCTCTGGGAAGCCCGCTTCCCCCTGCAGACAGGCG CTGTGACACGCTTGCGCCCCGGTC	303	aggaagagagATAAGATTT TTTGGGGAATTTGTTG	cagtaatacactcactat aggagaaggctAAAC CCTAAATTCAAAT CCTACCC
A4b_cg25333225_4	GAGCCAGAGACGAACCTCTGAGAGAGCGAGAGCGCAAGA GCAGGGCAGAGAAGGAGGCGGGCTCACAGCAGGGCGC CTGCGCACATTAGACAACCTTTGGGCCTGTGTGGGTTAGA GATGGGCCTCGCTAGGCCGTGGGGCGGAAAGGCGAGA TGGGGCTACGCAGGCGCACTAGGACTCCGACGTCTGCT TTCCCTAGGGCCTAGCCGTAGTGATTATTTCCCGGCAGC CCCTAGTCCGTATTTACATACGCTTTGATCTTGTTTACT	283	aggaagagagTAATTTTGG GTTTGTGTGGGTTAG	cagtaatacactcactat aggagaaggctAACT TCAAAACCTAAAC TCAACCAT

CAGTTCTGGGTTCTGACGCTCTTAAAGGACCCCGGAGC
 GGGAGTTAGTTAGGAGGGGTTTCCTTTTCAGTGTGCTCG
 TCATGGTTGAGCTCAGGCTCTGAAGCCACACAACCCAG
 ACTTGAATTCCTCCTTTCTGGTACTGTGGGCAGAAGG
 CTTGACCGTTTTCTGTGCCTCAGTATCCACATCTGCTAAA
 TGGAAACAATCATCGCTGTGTCGCTGGCTTTTGAACCAT
 TTTAATGTAAACAAAACGGTTTTTACCAATTAGTTCATTAC
 TGCAAGATATTAATGCCACCTGCTCTCATCGTACATATTC
 CAGTGAATCTGGGATATA

A5a_cg20587409_5	<p>TTCCTGTTTTTTAATATTAAGAAGAGACCTCTTTTCACA AATGATTGAGGAGGCTAGTAAAATGCAGTGAAAAGAGCA AGTTCTTTGGGAGTCAGACAGCATCGGATCCAGCCAGA GCGCAGCAGCACTAGTAGAGTAACTTAGATTAAGTCCTT AACTTCCAGTGCCTCTTTGTAAACTGGGAATAATAAT AATAATAAATAAATAGTGAATAATATTTATGCAAAGTACTT AGCACAGTGTCTAACACATAAAAGCTTAGGAGAAGTTGG CTATTAAGCATTCTAGGAACCCGAGGACTCAGGGTTGT TATTTTAGTTCTTGGAATCAAAGGAACTTGAAGCGTAGTC ATTCCTAAACTAAGTTAAACCAGAACTAGCGCTTTCAGCA ATGATGTAATTTGATCTCTTCGGGCCACCGTCGCTGAGC ATGCGCAAATAAACGTGGCGGGACGTATGTGTCATGGC GCTCTCCATCTAAAGTCTGTGCAGCTTCCGGAGAGTGGC GGGTTGATTTTCTCACTTTGGACTGGTTTTTACTTCCCGA CTTCTGGGTAAGGGTGGCCGATGGGTCCATGTGGGGCA GAGTGTGGCC</p>	338	<p>aggaagagagTTTAGGAGA AGTTGGTTATTAAGTAT TTT</p>	<p>cagtaatcgactcactat agggagaaggctAACC AACTCTACCCCA CATAAAC</p>
A6a_cg23951816_9	<p>GAATAGAGGGAGAGAAGGCAATTAGGGGGCTGTGTGGT TGGGGCTTTTTACTTGTCAAAGTGGAGGCTCATTGCTTTT GTTACCACCATGTAAAGGGAATGGCAGGGTTTTGTTTTG GTATTAAGGGCCTTAAAGATTGCTGCCTGACACTTACAC</p>	344	<p>aggaagagagTTTGTTTAG TTTTGTGTTTGTGTTGA</p>	<p>cagtaatcgactcactat agggagaaggctATCT TTTCCACCCCAA ATATTA</p>

AGGAAGGCCAGATAAGTCCAGAAAGAGAGGCCACGGAG
 CAGTATTTTAAAAAAGAAGAAAAGAAGAAATGAAGAATTC
 CCACTTTTTTGTATTTTTAAGCCTGCCAGCCCTTGTTT
 TTGTTGATTTTTTTGGTGGGGAGCGGCGAGGGTGATTAA
 CTTGTCCCGTGTCTTAACAGATTGTTTAAAGACCTACATC
 GTGAAACAGTCTTGTGGGCTTGGAGATAAACAGATGGG
 CAATCAGTTCAGTAAATTA AAAACATCTCTACCTTAGGTG
 TCCTGCCGCCAGAAGAAACCAAATCTTTGCCACTGGTCT
 TGCTTCAAGCCACACAAGTCCAGGATCCCATGTGGCAGC
 AGGGGTGTTTTCTTACCTGCATGCAAATCGGAGTTGTGT
 GTTTGCTTTTGAAGAGGGTGAAATGGCTTAACACTTGGG
 GTGGGAAAAGACC

Amplicon name consist of the following parts: Laboratory number_CpG site to replicate included in target_primer number. Red marked sequence is the target sequence of amplification. Left and right primer include Taq.

Table 2: Primer sequence for pyrosequencing.

Gene	Size	PCR primer forward	PCR primer reverse	Pyrosequencing primer
<i>TMEM57</i>	333	Biotine-GTGTGGGGTTTGTGTTTGTATT	AAACCCTATAAATATAACATCCTATATTAC	AAAAACATCCCAAAAAC
<i>CASZ1</i>	116	Biotine-TTTTTAAGGTGGTTAATAAGAAGGG	AAACTTCTCCTTTCCAAAAAATAAC	CATCTCTATCCCAAAAAT

Table 3: Annotation information of detected CpG sites.

CpG site	Location on chromosome	Annotated gene	Location within gene	Relation to CpG island
cg09154213	1:25756855	<i>TMEM57</i>	TSS1500	CpG Island
cg22800477	1:10853793	<i>CASZ1</i>	5'UTR	N Shore
cg18514820	10:17271944	<i>VIM</i>	Body	CpG Island
cg11057824	14:50471938	<i>C14orf182</i>	Body	S Shore
cg25333225	19:40791658	<i>AKT2</i>	TSS1500	S Shore
cg22876894	20:1783624	-	-	N Shore

Chromosome information is based on Genome Build 37. TSS: Transcription start site; CpG island: site with an accumulation of CpG sites; shore: located next to the CpG island; shelf: located more than 2 kb away from the CpG island; S/N superior of shore or shelf: specifies the location in relation the CpG island

Table 4: Characteristics of the study population for DNA methylation analysis (n=617) for 2-hour insulin.

	Median (25 th ; 75 th percentile)	%
Sex [% male]	-	48.30
Age [years]	68 (65; 72)	-
BMI [kg/m²]	27.55 (25.16; 30.31)	-
Waist circumference [cm]	96.10 (88.40; 103.50)	-
Fasting serum glucose [mmol/l]	5.39 (5.06; 5.72)	-
2-hour serum glucose [mmol/l]	6.50 (5.39; 7.78)	-
HbA1c [%]	5.60 (5.30; 5.80)	-
Glucose tolerance status [%]		
NGT	-	68.56
IFG	-	5.67
IGT	-	20.26
Combined IFG and IGT	-	5.51
Insulin [μU/ml]	4.70 (3.30; 7.50)	-
2-hour insulin [μU/ml]	50.20 (28.70; 78.10)	-
HOMA-IR	1.12 (0.76; 1.82)	-
C-reactive protein [mg/l]	1.37 (0.71; 2.53)	-
Leucocytes [/nl]	5.60 (4.80; 6.40)	-
Cholesterol [mmol/l]	5.79 (5.12; 6.51)	-
Triglycerides [mmol/l]	1.25 (0.91; 1.71)	-
Systolic blood pressure [mmHg]	126.50 (114.50; 138.00)	-
Diastolic blood pressure [mmHg]	74.50 (68.50; 81.50)	-
Alcohol consumption [g/day]	7.64 (0.00; 20.00)	-
Smoking status [%]		
never	-	52.84
ex	-	39.22
current	-	7.95
Physically active [%] (combination of activity during summer and winter)	-	56.89
CD8⁺ T cells [%]⁺	0.09 (0.05; 0.15)	-
CD4⁺ T cells [%]⁺	0.15 (0.11; 0.19)	-
Natural killer cells [%]⁺	0.02 (0.00; 0.04)	-
B cells [%]⁺	0.04 (0.03; 0.06)	-
Monocytes [%]⁺	0.11 (0.10; 0.14)	-
Granulocytes [%]⁺	0.64 (0.58; 0.70)	-

+ Data are presented for the estimated white blood cell proportion using a recently published method [142]

NGT: normal glucose tolerance; IFG: impaired fasting glucose; IGT: impaired glucose tolerance

Table 5: Characteristics of the study population for DNA methylation analysis (n=1,440) HOMA-IR.

	Median (25 th ; 75 th percentile)	%
Sex [% male]	-	47.15
Age [years]	59 (53; 67)	-
BMI [kg/m²]	27.06 (24.51; 30.05)	-
Waist circumference [cm]	93.55 (84.50; 102.33)	-
Fasting serum glucose [mmol/l]	5.28 (4.94; 5.61)	-
2-hour serum glucose [mmol/l]	6.00 (5.00; 7.17)	-
HbA1c [%]	5.50 (5.20; 5.70)	-
Glucose tolerance status [%]		
NGT	-	77.29
IFG	-	4.93
IGT	-	14.37
Combined IFG and IGT	-	3.40
Insulin [μU/ml]	4.10 (2.80; 6.70)	-
2-hour Insulin [μU/ml]	50.20 (28.70; 78.10)	-
HOMA-IR	0.97 (0.64; 1.59)	-
C-reactive protein [mg/l]	1.14 (0.58; 2.23)	-
Leucocytes [/nl]	5.50 (4.70; 6.50)	-
Cholesterol [mmol/l]	5.71 (5.11; 6.43)	-
Triglycerides [mmol/l]	1.22 (0.85; 1.70)	-
Systolic blood pressure [mmHg]	122.20 (111.00; 134.50)	-
Diastolic blood pressure [mmHg]	75.50 (69.50; 82.50)	-
Alcohol consumption [g/day]	8.57 (0.00; 22.86)	-
Smoking status [%]		
never	-	44.93
ex	-	40.00
current	-	15.07
Physically active [%] (combination of activity during summer and winter)	-	60.12
CD8⁺ T cells [%]⁺	0.09 (0.05; 0.14)	-
CD4⁺ T cells [%]⁺	0.16 (0.12; 0.21)	-
Natural killer cells [%]⁺	0.02 (0.002; 0.04)	-
B cells [%]⁺	0.05 (0.03; 0.06)	-
Monocytes [%]⁺	0.12 (0.10; 0.13)	-
Granulocytes [%]⁺	0.63 (0.57; 0.69)	-

+ Data are presented for the estimated white blood cell proportion using a recently published method [142]

NGT: normal glucose tolerance; IFG: impaired fasting glucose; IGT: impaired glucose tolerance

Table 6: Annotation information of detected CpG sites.

CpG site	Location on chromosome	Annotated gene	Location within gene	Relation to CpG island
cg00574958	11:68607622	<i>CPT1A</i>	5'UTR	N Shore
cg06500161	21:43656587	<i>ABCG1</i>	gene body	S Shore
cg07504977	10:102131012	unannotated	-	N Shelf
cg11024682	17:17730094	<i>SREBF1</i>	gene body	S Shelf
cg22040809	6:26522578	<i>HCG11</i>	gene body	CpG Island
cg07092212	11:46382544	<i>DGKZ</i>	gene body	-
cg09613192	2:181388538	unannotated	-	-
cg09694782	2:97408799	unannotated	-	S Shelf
cg11376147	11:57261198	<i>SLC43A1</i>	gene body	-
cg17266233	11:46382725	<i>DGKZ</i>	gene body, TSS1500 *	-
cg17971578	1:36852463	<i>STK40</i>	TSS1500	S Shore
cg22065733	8:11801320	unannotated	-	-
cg04161365	17:27230393	<i>DHRS13</i>	TSS1500	S Shore
cg13016916	7:137660322	<i>CREB3L2</i>	gene body	-
cg20477259	6:31544960	<i>TNF</i>	gene body	N Shelf

Chromosome information is based on Genome Build 37. TSS: Transcription start site; * depending and transcription variant; CpG island: site with an accumulation of CpG sites; shore: located next to the CpG island; shelf: located more than 2kb away from the CpG island; S/N superior of shore or shelf: specifies the location in relation the CpG island

Table 7: Association between DNA methylation at cg09694782/cg13016916 and different phenotypes based on quintiles of methylation level.

a

	1.Quintile (n=290)	2.Quintile (n=289)	3. Quintile (n=289)	4. Quintile (n=289)	5.Quintile (n=290)	
Phenotype	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	p for trend (bonf. adjusted)
Age [years] #	61.29 (8.6)	59.54 (8.65)	59.69 (8.66)	59.81 (9.13)	58.92 (8.47)	0.044
BMI [kg/m ²] #	27.59 (4.49)	27.93 (4.6)	27.5 (4.41)	27.53 (4.26)	27.04 (3.96)	0.917
Waist circumference [cm]	94.21 (13.29)	94.38 (13.4)	93.92 (13.32)	92.75 (12.07)	92.88 (12.34)	1
Fasting glucose [mmol/l] #	5.38 (0.55)	5.28 (0.5)	5.35 (0.53)	5.24 (0.49)	5.29 (0.55)	0.301
2-hour glucose [mmol/l] #	6.45 (1.77)	6.17 (1.73)	6.3 (1.72)	6.03 (1.63)	6.14 (1.67)	0.259
HbA1c [%]	5.48 (0.32)	5.47 (0.35)	5.49 (0.32)	5.45 (0.31)	5.46 (0.3)	1
C-reactive protein [mg/l] #	1.89 (1.88)	1.66 (1.52)	1.69 (1.54)	1.83 (1.73)	1.56 (1.61)	1
Fasting insulin [μIU/ml] #	7.09 (6.96)	6.67 (6.86)	6.58 (7.46)	5.77 (6.41)	5.26 (5.88)	3.67x10 ⁻³
2-hour insulin [μIU/ml] #	68.83 (54.4)	65.74 (49.27)	62.86 (47.83)	54.59 (37.43)	57.62 (59.59)	0.261
HOMA-IR #	1.75 (1.84)	1.59 (1.71)	1.63 (1.99)	1.38 (1.61)	1.26 (1.53)	2.91x10 ⁻³
Cholesterol [mmol/l] #	5.77 (1.08)	5.83 (1.03)	5.84 (1.01)	5.79 (0.98)	5.76 (0.91)	1
Triglycerides [mmol/l] #	1.38 (0.81)	1.49 (0.92)	1.53 (1.29)	1.4 (0.97)	1.43 (0.97)	1
Systolic blood pressure [mm Hg]	123.21 (20.53)	123.27 (17.43)	124.1 (18.22)	122.17 (18.17)	123.74 (16.73)	1
Diastolic blood pressure [mm Hg]	76.14 (10.78)	76.5 (9.41)	76.67 (9.94)	75.4 (9.59)	76.38 (9.61)	1
CD8 ⁺ T cells #+	0.11 (0.07)	0.1 (0.07)	0.1 (0.07)	0.1 (0.07)	0.11 (0.07)	2.92x10 ⁻⁷³
CD4 ⁺ T cells +	0.16 (0.06)	0.16 (0.06)	0.16 (0.06)	0.16 (0.06)	0.16 (0.06)	2.66x10 ⁻¹¹
Natural killer cells #+	0.02 (0.02)	0.02 (0.02)	0.03 (0.03)	0.02 (0.02)	0.02 (0.02)	1.54x10 ⁻⁸
B cells #+	0.05 (0.03)	0.05 (0.02)	0.05 (0.02)	0.05 (0.04)	0.05 (0.02)	1.59x10 ⁻⁵
Monocytes +	0.12 (0.02)	0.12 (0.03)	0.12 (0.02)	0.12 (0.03)	0.12 (0.03)	1.32x10 ⁻⁶
Granulocytes +	0.63 (0.08)	0.63 (0.09)	0.64 (0.09)	0.64 (0.1)	0.63 (0.09)	1.74x10 ⁻¹¹

	1. Quintile (n=290)		2. Quintile (n=289)		3. Quintile (n=289)		4. Quintile (n=289)		5. Quintile (n=290)	
	number	number	p-value	number	p-value	number	p-value	number	p-value	
Sex [male/female]	147/143	128/161	0.133	141/148	0.678	119/170	0.024	146/144	0.967	
Glucose status [combination of IFG and IGT / IFG / IGT / normal]	14/21/52/203	8/13/39/229	0.083	12/12/47/218	0.336	5/10/31/243	5.0x10 ⁻⁴ *	10/16/39/225	0.244	

b

	1.Quintile (n=290)	2.Quintile (n=289)	3. Quintile (n=289)	4. Quintile (n=289)	5.Quintile (n=290)	p for trend (bonf. adjusted)
Phenotype	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	
Age [years] #	59.59 (8.77)	60.08 (8.64)	60.13 (8.67)	59.34 (8.57)	60.29 (9.02)	1
BMI [kg/m²] #	27.46 (4.19)	27.66 (4.4)	27.3 (4.31)	27.51 (4.47)	27.52 (4.39)	1
Waist circumference [cm]	93.47 (12.53)	94.51 (13.35)	93.11 (12.6)	93.64 (13.76)	93.3 (12.4)	1
Fasting glucose [mmol/l] #	5.28 (0.54)	5.29 (0.54)	5.32 (0.53)	5.35 (0.51)	5.32 (0.51)	1
2-hour glucose [mmol/l] #	6.09 (1.63)	6.29 (1.77)	6.14 (1.65)	6.34 (1.84)	6.23 (1.65)	1
HbA1c [%]	5.47 (0.31)	5.47 (0.33)	5.44 (0.31)	5.48 (0.34)	5.48 (0.32)	1
C-reactive protein [mg/l] #	1.78 (1.74)	1.77 (1.6)	1.56 (1.62)	1.72 (1.69)	1.79 (1.7)	1
Fasting insulin [μU/ml] #	5.27 (5.75)	6.21 (6.26)	6.15 (6.55)	7.19 (8.12)	6.49 (6.77)	0.128
2-hour insulin [μU/ml] #	60.62 (56.55)	56.97 (38.65)	62.04 (45.97)	64.18 (53.26)	68.7 (56.18)	1
HOMA-IR #	1.26 (1.57)	1.51 (1.6)	1.48 (1.65)	1.75 (2.04)	1.6 (1.82)	0.103
Cholesterol [mmol/l] #	5.78 (0.98)	5.85 (1.08)	5.82 (1.05)	5.74 (0.94)	5.8 (0.98)	1
Triglycerides [mmol/l] #	1.4 (0.87)	1.49 (1.01)	1.5 (1.33)	1.43 (0.9)	1.4 (0.85)	1
Systolic blood pressure [mm Hg]	123.96 (17.7)	122.83 (18)	122.88 (17.71)	123.43 (18.1)	123.65 (19.75)	1
Diastolic blood pressure [mm Hg]	76.28 (9.83)	76.06 (10.06)	75.93 (9.43)	76.74 (9.86)	76.15 (10.28)	1

CD8⁺ T cells #+	0.1 (0.07)	0.1 (0.07)	0.1 (0.06)	0.1 (0.06)	0.1 (0.07)	1
CD4⁺ T cells +	0.16 (0.06)	0.16 (0.06)	0.17 (0.06)	0.17 (0.06)	0.17 (0.06)	1
Natural killer cells #+	0.03 (0.03)	0.03 (0.03)	0.02 (0.02)	0.03 (0.03)	0.02 (0.02)	0.396
B cells #+	0.05 (0.04)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.494
Monocytes +	0.12 (0.02)	0.12 (0.02)	0.12 (0.02)	0.12 (0.02)	0.12 (0.03)	1
Granulocytes +	0.63 (0.1)	0.64 (0.08)	0.63 (0.09)	0.62 (0.08)	0.63 (0.08)	1

	1. Quintile (n=290)		2. Quintile (n=289)		3. Quintile (n=290)		4. Quintile (n=289)		5. Quintile (n=290)	
	number	number	p-value	number	p-value	number	p-value	number	p-value	
sex [male/female]	138/148	140/146	0.702	131/155	0.827	138/147	0.766	129/157	0.698	
glucose status [combination of IFG and IGT / IFG / IGT / normal]	8/12/35/231	11/17/43/215	0.347	10/19/40/217	0.394	9/12/50/215	0.274	11/11/39/225	0.764	

Means, standard deviations and p-values for trend (Bonferroni corrected) are presented for the different quintiles for the continuous phenotypes, and total numbers in the different quintiles and p-values (comparing the quintile vs 1st quintile) for categorical variables. * p-value was significant after correction for multiple testing. # variables were log transformed for determination of p-values); + cell types were estimated using method developed by Houseman et al. [142], IFG: impaired fasting glucose, IGT: impaired glucose tolerance; (a) cg09694782 (one sample without methylation value for this CpG site) and (b) cg13016916 (18 samples without methylation value for this CpG site)

Table 8: Pathway analysis based on the top 1,000 CpG sites associated with 2-hour glucose (a), fasting insulin (b) and HOMA-IR (c) using Ingenuity.

a

Adjusted for age, sex, and estimated white blood cell proportions (model 1)			Adjusted for age, sex, estimated white blood cell proportions, and BMI (model 2)		
Ingenuity Canonical Pathways	B-H-adjusted p-value	Ratio	Ingenuity Canonical Pathways	B-H-adjusted p-value	Ratio
<u>Growth Hormone Signaling</u>	0.022	11/69	<u>Thyroid Hormone Metabolism II (via Conjugation and/or Degradation)</u>	0.018	7/26
<u>Role of NFAT in Cardiac Hypertrophy</u>	0.028	18/176	Serotonin Degradation	0.163	8/52
<u>Thyroid Hormone Metabolism II (via Conjugation and/or Degradation)</u>	0.035	6/26	Nicotine Degradation III	0.186	7/44
<u>α-Adrenergic Signaling</u>	0.035	11/85	Melatonin Degradation I	0.198	7/47
<u>Protein Kinase A Signaling</u>	0.035	28/368	Ephrin Receptor Signaling	0.198	15/172
<u>CREB Signaling in Neurons</u>	0.048	16/169	Nicotine Degradation II	0.198	7/50
<u>CCR5 Signaling in Macrophages</u>	0.048	9/67	Superpathway of Melatonin Degradation	0.213	7/52
<u>G Protein Signaling Mediated by Tubby</u>	0.048	6/31	Growth Hormone Signaling	0.254	8/69
<u>Leukocyte Extravasation Signaling</u>	0.050	17/193	G Protein Signaling Mediated by Tubby	0.258	5/31
<u>Sperm Motility</u>	0.050	12/113	Factors Promoting Cardiogenesis in Vertebrates	0.258	9/89

b

Adjusted for age, sex, and estimated white blood cell proportions (model 1)			Adjusted for age, sex, estimated white blood cell proportions, and BMI (model 2)		
Ingenuity Canonical Pathways	B-H-adjusted p-value	Ratio	Ingenuity Canonical Pathways	B-H-adjusted p-value	Ratio
cAMP-mediated signaling	0.511	16/216	Netrin Signaling	0.109	7/39
Hypoxia Signaling in the Cardiovascular System	0.511	7/63	Role of Tissue Factor in Cancer	0.109	12/107
B Cell Receptor Signaling	0.511	13/171	Thrombin Signaling	0.136	16/187
PAK Signaling	0.511	8/88	Huntington's Disease Signaling	0.136	18/226
NGF Signaling	0.511	9/106	Tec Kinase Signaling	0.136	14/156
Role of IL-17F in Allergic Inflammatory Airway Diseases	0.511	5/41	SAPK/JNK Signaling	0.136	10/93
Role of Tissue Factor in Cancer	0.511	9/107	Reelin Signaling in Neurons	0.136	9/79
Superpathway of Inositol Phosphate Compounds	0.511	13/183	Glutathione Redox Reactions I	0.149	4/17
Intrinsic Prothrombin Activation Pathway	0.511	4/28	Axonal Guidance Signaling	0.162	27/425
Glioblastoma Multiforme Signaling	0.511	11/145	B Cell Receptor Signaling	0.162	14/171

c

Adjusted for age, sex, and estimated white blood cell proportions (model 1)			Adjusted for age, sex, estimated white blood cell proportions, and BMI (model 2)		
Ingenuity Canonical Pathways	B-H-adjusted p-value	Ratio	Ingenuity Canonical Pathways	B-H-adjusted p-value	Ratio
Hypoxia Signaling in the Cardiovascular System	0.353	8/63	Ephrin Receptor Signaling	0.778	17/172
Reelin Signaling in Neurons	0.353	9/79	Actin Cytoskeleton Signaling	0.964	18/210
IL-15 Signaling	0.353	8/66	ILK Signaling	0.964	16/181
TR/RXR Activation	0.353	9/85	ATM Signaling	0.964	8/59
FAK Signaling	0.353	9/86	B Cell Receptor Signaling	0.964	15/171
PAK Signaling	0.353	9/88	Role of Tissue Factor in Cancer	0.964	11/107
Role of Tissue Factor in Cancer	0.366	10/107	Hypoxia Signaling in the Cardiovascular System	0.964	8/63
Estrogen-Dependent Breast Cancer Signaling	0.393	7/62	SAPK/JNK Signaling	0.964	10/93
Axonal Guidance Signaling	0.398	25/425	Reelin Signaling in Neurons	0.964	9/79
FLT3 Signaling in Hematopoietic Progenitor Cells	0.398	7/71	Agrin Interactions at Neuromuscular Junction	0.964	8/67

a) 2-hour glucose, b) fasting insulin, c) HOMA-IR

The p-value corrected using the Benjamini-Hochberg method for multiple testing is presented for each pathway, besides the ratio (number of genes uploaded in the software/total number of genes included in the pathway). Underlined pathways are significant after correction for multiple testing using Benjamini-Hochberg.

Table 9: Summary of the analysis for association between DNA methylation and gene expression showing the top association per CpG sites as long as p-value < 0.05.

CpG site	Transcript	Coefficient	p-value	Adj. p-value	Annotated gene for CpG site	Annotated gene for transcript
cg06500161	ILMN_2329927	-3.623	2.5×10^{-12}	1.1×10^{-9}	<i>ABCG1</i>	<i>ABCG1</i>
cg00574958	ILMN_1744835	-3.423	4.1×10^{-3}	0.19	<i>CPT1A</i>	<i>MRPL21</i>
cg22040809	ILMN_1700067	-1.175	2.3×10^{-3}	0.15	<i>HCG11</i>	<i>BTN3A2</i>
cg09694782	ILMN_1769752	0.905	8.0×10^{-4}	0.10	unannotated	<i>LOC90342</i>
cg11376147	ILMN_2233099	1.354	6.8×10^{-3}	0.28	<i>SLC43A1</i>	<i>SSRP1</i>
cg17266233	ILMN_1660882	-1.213	0.03	0.62	<i>DGKZ</i>	<i>CHRM4</i>
cg04161365	ILMN_1716441	0.938	0.03	0.59	<i>DHRS13</i>	<i>SNORD4A</i>
cg20477259	ILMN_2150787	-1.958	9.1×10^{-4}	0.10	<i>TNF</i>	<i>HLA-C</i>
cg11024682	ILMN_2060770	-1.021	7.9×10^{-3}	0.30	<i>SREBF1</i>	<i>RAI1</i>
cg09613192	ILMN_2390338	-0.658	1.5×10^{-3}	0.11	unannotated	<i>UBE2E3</i>
cg22065733	ILMN_2144088	-3.103	1.4×10^{-3}	0.11	unannotated	<i>FDFT1</i>

7. List of Tables

Table 1: Cycling protocol for amelogenin PCR.

Table 2: Cycling protocol for Infinium HumanMethylation450 BeadChip respectively EpiTYPER® process.

Table 3: Cycling protocol for PCR amplification for EpiTYPER® process.

Table 4: Cycling protocol for PCR amplification for pyrosequencing process.

Table 5: Number of probes and samples excluded during preprocessing.

Table 6: Regression models for methylome-wide association study by incident T2D.

Table 7: Regression models for replication of the methylome-wide association study by incident T2D (LOLIPOP study).

Table 8: Regression models for methylome-wide association study by measures of glucose metabolism.

Table 9: Study characteristics of the discovery study – a total of 196 matched case-controls pairs.

Table 10: Study characteristics of the replication study – a total of 395 matched case-controls pairs.

Table 11: Results of methylome-wide association analysis with incident T2D for the discovery study.

Table 12: Results of methylome-wide analysis with T2D for the replication study showing the leading CpG sites from the discovery study.

Table 13: Results of methylome-wide analysis with incident T2D for the replication study showing the significant CpG sites.

Table 14: Results of the pathway analyses of the discovery study without adjustment presenting the first ten canonical pathways.

Table 15: Results for pathway analysis of the discovery study after adjustment for BMI presenting the first ten canonical pathways.

Table 16: Replication testing for association with incident T2D in Europeans.

Table 17: Relative risk for incident T2D for Europeans (replication).

Table 18: Study characteristics of the study population for DNA methylation analysis (n=1,448) for fasting glucose, 2-hour glucose, fasting insulin, and HbA1c.

Table 19: Results for association of genome-wide DNA methylation levels and fasting glucose, 2-hour glucose, fasting insulin, and HOMA-IR as well as 2-hour insulin with reduced number of CpG sites after adjustment for different potential confounders.

Table 20: Association between DNA methylation at cg06500161 and different phenotypes based on quintiles of methylation level.

Table 21: Pathway analysis based on the top 1,000 CpG sites associated with fasting glucose using Ingenuity.

Table 22: Characteristics of the study population for gene expression analysis (n=533).

Table 23: Summary of the analysis for association between DNA methylation and gene expression, as well as between DNA methylation and phenotypes.

Table 24: General function of proteins encoded by annotated genes of genome-wide significant CpG sites associated with incident T2D in humans.

Table 25: General function of proteins encoded by annotated genes of replicated genome-wide significant CpG sites associated with incident T2D in Indian Asians and Europeans.

Table 26: General function of proteins encoded by annotated genes of genome-wide significant CpG sites associated with glycemic parameters in humans.

Appendix table 1: Primer sequence for EpiTYPER®.

Appendix table 2: Primer sequence for pyrosequencing.

Appendix table 3: Annotation information of detected CpG sites.

Appendix table 4: Characteristics of the study population for DNA methylation analysis (n=617) for 2-hour insulin.

Appendix table 5: Characteristics of the study population for DNA methylation analysis (n=1,440) HOMA-IR.

Appendix table 6: Annotation information of detected CpG sites.

Appendix table 7: Association between DNA methylation at cg09694782/cg13016916 and different phenotypes based on quintiles of methylation level.

Appendix table 8: Pathway analysis based on the top 1,000 CpG sites associated with 2-hour glucose (a), fasting insulin (b) and HOMA-IR (c) using Ingenuity.

Appendix table 9: Summary of the analysis for association between DNA methylation and gene expression showing the top association per CpG sites as long as p-value < 0.05.

8. List of Figures

Figure 1: Pathogenesis of T2D.

Figure 2: Identified T2D susceptible loci divided into year of discovery.

Figure 3: Contribution of genetic and environmental factors to development of T2D and related traits and missing heritability.

Figure 4: Overview of epigenetic mechanisms.

Figure 5: Overview of different inheritance possibilities.

Figure 6: Interaction of genetic and epigenetic factors.

Figure 7: Definition of cases and controls for nested case-control study.

Figure 8: Flow diagram for study population of nested case-control study.

Figure 9: Flow diagram for study population of cross-sectional study.

Figure 10: Principle of bisulfite conversion.

Figure 11: Chemical basis of conversion of cytosine to uracil.

Figure 12: Overview of Illumina laboratory process.

Figure 13: Schematic presentation of the two different chemistry technologies used at the Infinium HumanMethylation450 BeadChip.

Figure 14: Overview of EpiTYPER[®] workflow.

Figure 15: Overview of pyrosequencing workflow.

Figure 16: Results of methylome-wide association analysis with incident T2D.

Figure 17: Differences for methylation degree of methylome-wide CpG sites between cases and controls.

Figure 18: Positions of CpG sites a) cg11057824 (annotated to *C14orf182*) and b) cg25333225 (annotated to *AKT2*) within the amplicon, respectively.

Figure 19: Results of methylome-wide association analysis with incident T2D.

Figure 20: Genome-wide associations between methylation and (a) fasting glucose level, (b) 2-hour glucose level, (c) fasting insulin level, (d) HOMA-IR after adjustment for sex, age, and estimated white blood cell proportions (model 1).

Figure 21: Distribution of BMI (a), fasting insulin (b), 2-hour insulin (c), and HOMA-IR (d) (y-axis) for the different degrees of methylation presented as quintiles (x-axis).

Appendix figure 1: Line plots of results comparing 750 ng and 1,000 ng DNA.

Appendix figure 2: Screen shot from the control dashboard.

9. Description of own contribution

This doctoral thesis comprises of two different projects with different kinds of analysis, where some of the presented results are contributed by others. Therefore in the following chapter my own contribution to the project is described in more detail.

I performed the analysis of association of DNA methylation and incident T2D the laboratory work for the HumanMethylation450 BeadChip on my own, as well as planning EpiTYPER[®] and pyrosequencing experiments. Further, I carried out the statistical analysis of the combined analysis of EpiTYPER[®] and pyrosequencing results as well as the interpretation of all results and writing of the manuscript.

Moreover, I carried out the analysis of the association between DNA methylation and measures of glucose metabolism, planning and performing all analysis, excluding the generation of DNA methylation data. Furthermore, I interpreted data and wrote the manuscript for the publications “Whole-blood DNA methylation patterns and measures of glucose metabolism in the KORA F4 study” and “Association of genome-wide whole blood methylome pattern and incident diabetes in the KORA study” (see chapter 10.1.).

10. List of publications and presentations

10.1. Publications included in this thesis

Original publication:

Chambers JC*, Loh M*, Lehne B*, Drong A*, Kriebel J*, Motta V*, Rota F, Scott WR, Zhang W, Tan ST, Campanella G, Chadeau-Hyam M, Adamowicz-Brice M, Afzal U, Bozaoglu K, Tarantini L, Wahl S, Abbott J, Ala-Korpela M, Albetti B, Bertazzi PA, Blancher C, Danesh J, de Lusignan S, Gieger C, Herder C, Jha S, Jones S, Jowett J, Kangas AJ, Kasturiratne A, Kato N, Kotea N, Kowlessur S, Pitkaniemi J, Punjabi P, Saleheen D, Soininen P, Tai ES, Thorand B, Tuomilehto J, Wickremasinghe AR, Zimmet P, Kyrtopoulos SA, Vineis P, Aitman TJ, Illig T, Grallert H, Scott J*, Jarvelin MR*, Bollati V*, Elliott P*, McCarthy MI*, Kooner JS*: DNA methylation markers in peripheral blood predict incident Type-2 diabetes amongst Indian Asians and Europeans. Submitted

* Contributed equally

Kriebel J*, Herder C*, Rathmann W*, Wahl S, Kunze S, Molnos S, Schramm K, Carstensen-Kirberg M, Waldenberger M, Gieger C, Peters A, Illig T, Prokisch H*, Roden M*, Grallert H*. Whole-blood DNA methylation patterns and measures of glucose metabolism in the KORA F4 study. Submitted

* Contributed equally

Kriebel J*, Wahl S*, Kunze S, Waldenberger M, Tost J, Busato F, Baurecht HJ, Thorand B, Meisinger C, Rathmann W, Gieger C, König W, Peters A, Schneider A, Flaquer A, Rospleszcz S, Strauch K, Roden M, Herder C, Grallert H, Illig T. Association of genome-wide whole blood methylome pattern and incident diabetes in the KORA study. In Preparation

* Contributed equally

10.2. Publications not included in this thesis

During doctoral thesis

Reviews:

Kriebel J, Illig T, Grallert H: Epigenetische Prozesse beim Typ-2-Diabetes. Diabetologe. 2013/3

Kriebel J, Illig T, Grallert H: Genomweite Assoziationsstudien (GWAS) – Möglichkeiten und Grenzen. Biospektrum. 2012/5

Kriebel J, Grallert H, Illig T: Typ-2-Diabetes-assoziierte Gene. Diabetologe. 2012/1

Original publication:

von Loeffelholz C, Lock JF, Döcke S, Birkenfeld AL, Lieske S, Hoppe S, **Kriebel J**, de las Heras Gala T, Grallert H, Raschzok N, Sauer IM, Weickert MO, Osterhoff MA, Horn P, Heller R, Bauer M, Jahreis G, Stockmann M, Pfeiffer AF: Palmitate Induces X-Box binding-Protein-1 and Drives Lipogenic Enzyme Expression in Human Nonalcoholic Fatty Liver Disease. J Hepatol. Submitted

Mahajan A, Sim X, Ng HJ, on behalf of the T2D-GENES consortium and **GoT2D consortium**: Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. PLoS Genetics. 2014. Accepted

Flaquer A, Baumbach C, Ladwig KH, **Kriebel J**, Waldenberger M, Grallert H, Baumert J, Meitinger T, Kruse J, Peters A, Emeny R, Strauch K: Mitochondrial genetic variants identified to be associated with posttraumatic stress disorder. Translational Psychiatry. 2015. Accepted

Tönjes A, Scholz M, Breitfeld J, Marzi C, Grallert H, Gross A, Ladenvall C, Schleinitz D, Krause K, Kirsten H, Laurila E, **Kriebel J**, Thorand B, Rathmann W, Groop L, Prokopenko I, Isomaa B, Beutner F, Kratzsch J, Thiery J, Fasshauer M, Klötting N, Gieger C, Blüher M, Stumvoll M, Kovacs P: Genome Wide Meta-analysis Highlights the Role of Genetic Variation in RARRES2 in the Regulation of Circulating Serum Chemerin. PLoS Genet. 2014 Dec 18;10(12):e1004854

Majithia AR, Flannick J, Shahinian P, Guo M, Bray MA, Fontanillas P, Gabriel SB; **GoT2D Consortium**; NHGRI JHS/FHS Allelic Spectrum Project, SIGMA T2D Consortium; T2D-GENES Consortium, Rosen ED, Altshuler D: Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. Proc Natl Acad Sci USA. 2014 Aug 25; Epub

Flaquer A, Baumbach C, **Kriebel J**, Meitinger T, Peters A, Waldenberger M, Grallert H, Strauch K: Mitochondrial Genetic Variants Identified to Be Associated with BMI in Adults. PLoS One. 2014 Aug 25;9 (8):e105116

Wang SR, Agarwala V, Flannick J, Chiang CW, Altshuler D; **GoT2D Consortium**, Hirschhorn JN: Simulation of Finnish population history, guided by empirical genetic data, to assess power of rare-variant tests in Finland. Am J Hum Genet. 2014 May 1;94 (5):710-20

Prior to doctoral thesis

Märker T, **Kriebel J**, Wohlrab U, Burkart V, Habich C: Adipocytes from New Zealand obese mice exhibit aberrant proinflammatory reactivity to the stress signal heat shock protein 60. *J Diabetes Res.* 2014; 2014:187153. doi: 10.1155/2014/187153. Epub 2014 Feb 5

Märker T, Sell H, Zillessen P, Glöde A, **Kriebel J**, Ouwens DM, Pattyn P, Ruige J, Famulla S, Roden M, Eckel J, Habich C: Heat shock protein 60 as a mediator of adipose tissue inflammation and insulin resistance. *Diabetes.* 2012 Mar;61 (3):615-25.

Märker T, **Kriebel J**, Wohlrab U, Habich C: Heat shock protein 60 and adipocytes: characterization of a ligand-receptor interaction. *Biochem Biophys Res Commun.* 2010 Jan 22;391 (4):1634-40

Gülden E, Märker T, **Kriebel J**, Kolb-Bachofen V, Burkart V, Habich C: Heat shock protein 60: evidence for receptor-mediated induction of proinflammatory mediators during adipocyte differentiation. *FEBS Lett.* 2009 Sep 3;583(17):2877-81

10.3. Poster presentation with results of this doctoral thesis

Kriebel J, Wahl S, Zeilinger S, Schramm K, Rathmann W, Roden M, Peters A, Illig T, Waldenberger M, Prokisch H, Grallert H, Herder C. Whole-blood DNA methylation patterns and glycemic traits in the KORA F4 study. American Society of Human Genetics (USA), 18.-22.October 2014/congress Epigenomics of common diseases. Cambridge (UK), 28.-31.October 2014

Kriebel J, Chambers JC, Loh M, Lehne B, Drong A, Motta V, Wahl S, Thorand B, Roden M, Herder C, Grallert H, Scott J, Jarvelin MR, Bollati V, Elliot P, McCarthy M, Kooner J. DNA methylation markers in peripheral blood predict incident type 2 diabetes amongst Asians and Europeans. SAB-Meeting, 23.-24.January 2014

Kriebel J, Wahl S, Zeilinger S, Rathmann W, Roden M, Peters A, Illig T, Waldenberger M, Grallert H, Herder C. Association of genome-wide DNA methylation patterns and insulin levels in KORA F4. DZD Workshop. Tübingen (Germany), 2013/congress Epigenomics of common diseases. Cambridge (UK), 07.-10.November 2013

10.4. Poster presentations prior to doctoral thesis

Burkart V, **Kriebel J**, Märker T, Habich C. Inflammatory adipocyte activation by heat shock protein 60 involves MAP-kinase- and NF κ B-dependent signaling pathways. 46. annual meeting of European association for the study of diabetes. Stockholm (Sweden), 20.-24. September 2010

Kriebel J, Märker T, Burkart V, Habich C. Hsp60-induced release of inflammatory mediators by adipocytes of the New Zealand obese mice is mediated by members of the MAP-kinase family and NF κ B. 25. annual meeting of Federation of the International Danube. Klausenburg (Romania), 9.-11. September 2010

Burkart V, Glden E, Märker T, **Kriebel J**, Habich C. Heat shock protein 60 and Toll-like receptor 4: Important regulators of inflammatory adipocyte activities. 70. Meeting of American Diabetes Association, Orlando (USA), 25.-29. June 2010

Kriebel J, Märker T, Burkart V, Habich C. „Die Hitzschockprotein 60-induzierte Freisetzung von Entzndungsmediatoren in Adipozyten der New Zealand obese Maus wird durch Mitglieder der MAP-Kinase Familie und NF κ B mediiert“, 45. Meeting of German Diabetes-Assoziation, Stuttgart (Germany), 12.-15. May 2010

11. Acknowledgement (Danksagung)

Ich möchte mich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben, sei es direkt oder indirekt.

Als erstes möchte ich mich bei meinem Doktorvater Herrn Prof. Dr. Thomas Illig für die Übernahme der Betreuung an der LMU und der Ermöglichung dieser Arbeit herzlich bedanken.

Ebenso gilt mein besonderer Dank Dr. Harald Gallert für jegliche Unterstützung und Motivation in der ganzen Zeit, sowie der Hilfe bei der Erstellung der Arbeit.

PD Dr. Christian Herder danke ich vielmals für die Bereitschaft der Teilnahme an meinem „Thesis Committee“ und den damit verbundenen Hilfestellungen und Ratschlägen.

Während der Zeit meiner Doktorarbeit zeigten alle drei großes Interesse und Engagement an meiner Arbeit und deren Fortschritt. Ebenso hatten Sie ein offenes Ohr für Probleme bzw. Fragen von meiner Seite und standen mir mit Rat und Tat zur Seite.

Ich möchte mich bei allen Kollegen bedanken, die mich die ganze Zeit unterstützt und auch mal motiviert haben. Speziell möchte ich mich bei Frau Dr. Kunze, Frau Simone Wahl, Frau Dr. Carola Marzi, Frau Sophie Molnos, Frau Dr. Melanie Waldenberger, Frau Dr. Eva Reischl, Frau Dr. Anja Kretschmer sowie Frau Liliane Pfeiffer für die wissenschaftlichen Diskussionen, konstruktive Kritik und Hilfestellungen/Anregungen bedanken. Frau Nicole Spada, Frau Nadine Lindemann, Frau Franziska Scharl und Frau Viola Maag möchte ich für ihre Unterstützung und Hilfe im Labor bedanken. Ebenso gilt mein Dank Frau Natascha Strauß, Rory Wilson sowie Herrn Dr. Fabio Spada für ihre Unterstützung.

Ebenso gilt mein Dank Andrea Schneider, Dr. Joerg Tost, Florence Busato, Hans-Jörg Baurecht, Dr. Antonia Flaquer, Susanne Rospleszcz und Dr. Maren Carstensen-Kirberg, die mit ihrer fachlichen Unterstützung ebenfalls zum Gelingen meiner Doktorarbeit beigetragen haben.

Ein ganz großer Dank gilt auch meiner Familie und meinen Freunden, die mich emotional wo immer es ging unterstützt haben und mir Mut gemacht haben, wenn es mal nicht so gut lief. Leider kann ich euch gar nicht alle persönlich erwähnen, aber ich weiß eure Hilfe zu schätzen und bin euch unendlich dankbar.

12. Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema „Epigenetic analysis of type 2 diabetes and measures of glucose metabolism“ selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum

Unterschrift Doktorandin