Association between plasma glucose in pregnancy and DNA methylation of candidate CpG sites:

A cross-sectional study

Nadezhda Kiryushchenko

Thesis submitted for the degree of
Master of Science in Molecular Bioscience
60 credits

Department of Bioscience
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

[May / 2019]

Association between plasma glucose in pregnancy and DNA methylation of candidate CpG sites: A cross-sectional study

Thesis submitted for the degree of Master of Science in Molecular Bioscience
60 credits

Department of Bioscience
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

[May / 2019]

Nadezhda Kiryushchenko

Year: 2019

Title: Association between plasma glucose in pregnancy and DNA methylation of candidate CpG sites

Author: Nadezhda Kiryushchenko

http://www.duo.uio.no/

Print: Reprosentralen, Universitetet i Oslo

Abstract

Background: Increased plasma glucose levels in pregnancy or gestational diabetes may have implications for future maternal health, such as increased risk of type 2 diabetes mellitus (T2DM) developing within 10 years. Increasing evidence indicate that epigenetics, such as DNA methylation, may play an important role in development of gestational diabetes. Epigenome-wide association studies (EWAS) explore association between traits and DNA methylation across the genome. The EPIPREG study (role of epigenetics in diabetes development during pregnancy) provided EWAS using the Illumina Methylation EPIC BeadChip (EPIC), and identified two CpG (cytosine – phosphate – guanine) sites associated with fasting plasma glucose (FPG) (cg08098128 and cg14120215) and one CpG site associated with 2 hour plasma glucose (2-h PG) (cg19327414) in Europeans.

Aims: The first aim of this project was to technically validate the EPIC using bisulfite pyrosequencing (BSP). Further aims were to explore how DNA methylation of top tree candidate CpG sites found from the EWAS in EPIPREG study was associated with different factors, and how DNA methylation of candidate CpG sites found in studies associated with T2DM was associated with FPG and 2-h PG.

Methods: The EPIPREG study provided EWAS in healthy pregnant Europeans (EU) (n=314) and South Asians (SA) (n=166) women, using the data collected from the STORK Groruddalen study. The BSP method was used for EPIC validation. The DNA methylation levels found in EPIPREG study and two other studies associated with T2DM, and other measurements collected at 28±2 gestational week were sorted according in tertiles according to low, medium and high methylation level at each CpG site.

Results: The strong correlation was observed ($r = 0.64$, $p < 0.001$) between EPIC and BSP methods. Tertiles of DNA methylation indicated a positive association between cg08098128 methylation and BMI ($p = 0.046$), and a negative association between cg14120515 methylation and age ($p < 0.001$) in Europeans, indicating that BMI and age may be confounding factors.

Four significant associations between CpG sites found in other studies and plasma glucose levels was observed (cg06500161 (SA, FPG), cg11024682 (SA, FPG), cg00574958 (EU, FPG), cg04999691 (EU, 2-h PG)), using ANOVA. The applying of Person correlation

indicated larger number of significant associations between CpG methylation and plasma glucose levels: cg11024682 (SA, FPG), cg07988171 (EU, FPG), cg07988171 (EU, 2-h PG), cg11024682 (SA, 2-h PG), cg18181703 (EU, 2-h PG), cg00574958 (EU, 2-h PG), cg19693031 (EU, FPG), cg06500161 (SA, FPG).

Conclusion: High technical validity of the EPIC was suggested due to strong correlation between EPIC and BSP. The association between BMI and age found in tertiles, probably identifies that BMI and age are confounding factors. Some of CpG sites found in other studies confirmed the association with plasma glucose levels in STORK Groruddalen study. Applying the Person correlation detected more associations.

Acknowledgements

This work was performed in the EPIPREG research group at Oslo University Hospital, Department of Endocrinology, Morbid Obesity and Preventive Medicine, in collaboration with EpiGen department at Akershus University Hospital, Lørenskog.

First and foremost, I would like to thank my main supervisor Christine Sommer. I am so grateful for your considerate support and guidance throughout this project. Thank you for always having time for my questions and explaining a lot of study material about diabetes and statistics. Your never-ending enthusiasm, constructive critic and highly informative feedbacks were helpful during the writing process.

Warm thanks to my great co-supervisor Gunn-Helen Moen, a genius programmer who helped me a lot with R work. Thank you for taking your time to help, answering my questions and giving many useful comments, helping me think on my own. I also wish to express my gratitude to my co-supervisor Kåre I. Birkeland. Thank you for giving me irreplaceable knowledge about metabolism and diabetes, and your valuable advices during my work with master thesis. Thank your all for being the best supervisors ever!

I also need to thank other members of our Epigenetic group: Elisabeth Qvigstad and Julia Onsrud Opsahl for always being kind and generous in giving me support. My special thanks goes to May Britt Dahl and Aina Fossum Moen who taught me the pyrosequencing technique and provided a great guidance during the laboratory work.

Further, I would like to thank my dear mother. You have always believed in me and this belief helped me all the way. Thanks to all my friends for bringing laughter and joy into my life. Thanks to the cutest cat ever, my dear Rysia. You did not sleep with me at nights with exam preparation during the study process.

Last, but definitely not least, my fiancé Alexey. You have always convinced me that I can do more that I am actually doing. Thanks that you persuaded me to find a study program that will be really interesting for me and to apply to the university, and I ended up here, where I feel I belong.

Oslo, May 2019

Nadezhda Kiryushchenko

Abbreviations

2-h PG	2 hour Plasma Glucose
450K	HumanMethylation450 BeadChip
ANOVA	Analysis of Variance
APS	Adenosine 5′-phosphosulfate
BMI	Body Mass Index
BSP	Bisulfite Pyrosequencing
CpG site	Cytosine – phosphate – Guanine site
dsDNA	double-stranded DNA
EPIC	Illumina Methylation EPIC BeadChip
EPIPREG	Role of epigenetics in diabetes development during pregnancy
EWAS	Epigenome-Wide Association Study
FPG	Fasting Plasma Glucose
GDM	Gestational Diabetes Mellitus
HOMA-B	Homeostatic Model Assessment of β cell Function
HOMA-IR	Homeostatic Model Assessment of Insulin Resistance
IADPSG	International Association of Diabetes and Pregnancy Study Groups
IR	Insulin Resistance
MBDs	Methyl-CpG-Binding Domains
NurD	Nucleosome Remodeling Deacetylase
OGTT	Oral Glucose Tolerance Test

PCR	Polymerase chain reaction
PPi	Pyrophosphate
SAM	S-adenosyl-methionine
SDH	L-serine dehydratase/L-threonine deaminase
ssDNA	single-stranded DNA
T2DM	Type 2 Diabetes Mellitus
WHO	World Health Organization
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil

Table of content

1 Introduction 1
1.1 Gestational Diabetes Mellitus 2
1.2 Epigenetics 4
2 Aims 10
3 Methods 11
3.1 STORK Groruddalen and the EPIPREG study 11
3.2 Validation study performed by the master student 17
3.3 Statistics 24
3.4 Ethics 25
4 Results 26
4.1 Cohort characteristics 26
4.2 Technical validation 28
4.3 DNA methylation analysis 29
4.3.1 Candidate CpG sites from EPIPREG 29
4.3.1 Candidate CpG sites from other studies 33
5 Discussion 37
5.1 Methodological considerations 37
5.1.1 Study design 37
5.1.2 Internal validity 37
5.1.3 External validity 40
5.2 Main findings 41
5.2.1 Technical validation 41
5.2.2 Candidate CpG sites 41
5.2.1 Possible biological mechanisms 42
6 Conclusions 46
References 47
7 Appendix 55

1 Introduction

Maternal metabolism during pregnancy undergoes a number of changes. This adaptation is essential to ensure the fetus with nutrients required to normal growth and development, and to meet the mothers' increased energy demands (1). Glucose is the main energy source for the fetus, which crosses the placenta freely (2). It is therefore very important to regulate glucose metabolism adequately.

There are two major changes that affect glucose metabolism during pregnancy: The blood volume increases, which result in decreased fasting plasma glucose (FPG) levels during the first trimester (3). After the first trimester, the insulin resistance (IR), the condition with impaired insulin sensitivity, increases (4). The process is required for adequate fetus development. Since insulin does not act properly and doesn't inhibit as much triglyceride degrading enzymes, the mother will use more fat than carbohydrates as energy source, leaving more glucose as fuel to the growing fetus (5).

Insulin and insulin resistance

Insulin is a peptide hormone, which is produced by pancreatic β cells (6, p.237, 880). The hormone has an anabolic effect: It stimulates cellular uptake of nutrient and ensures their storage. Thereby it increases triglycerides production as well as protein and glycogen syntheses, but it also inhibits triglyceride degrading enzymes (7, p.238-240). One of the most important functions of the hormone is to help muscle cells, liver and fat tissues to absorb, digest and store glucose for later use (8).

IR is a condition where cells have decreased response to insulin. Increased IR is observed to some degree in all pregnancies (9, 10), in association with the changed maternal metabolism.

During the developing IR, pancreatic β cells maintain the physiological level of blood sugar by increasing insulin secretion, which in turn leads to relative hyperinsulinemia. In this condition, the blood sugar level will be within the range normal as long as β cells are able to compensate with higher insulin production, although IR can be very high. However, if β cells are not able to compensate, hyperglycemia will develop (11).

1.1 Gestational Diabetes Mellitus

Gestational diabetes mellitus (GDM) is defined as an endocrine disorder characterized by hyperglycemia first recognized during pregnancy (12, 13). Even though the glucose levels are normalized after pregnancy, the disorder can affect both mother and the offspring later in life (14). A family history of type 2 diabetes mellitus (T2DM) (15) as well as a past history of GDM, maternal age over 40 years and obesity (body mass index (BMI) $\geq$35 kg/m^2) are important risk factors for the disorder (16). Findings from Europe indicate that ethnic minorities in Europe, especially South Asians, have higher risk of GDM (17).The prevalence of GDM has been estimated at 24.1% in Europeans and 41.4% in South Asians (18). However, the range varies largely with criteria used and across populations (19). Most counties use different criteria to diagnose GDM. GDM criteria in Norway is now: FPG $\geq$ 5.3 mmol/L and 2 hour plasma glucose (2-h PG) $\geq$ 9.0 mmol/L (20), but diagnosis criteria for GDM are not unanimous.

Implications for the mother and the offspring

As mentioned before, GDM may have significant implications for mother and offspring: during birth and later in the life (14). Fifty percent of women with GDM diagnosis will develop T2DM within ten years (21). The risk of T2DM developing after GDM increase with factors that can cause repeated IR like obesity and unhealthy high calorie diet (22).

Hyperglycemia is known to cause vascular endothelial cells damage (23). Changes in blood vessels play a role in developing of late complications or complications observed in diabetic patients over time. Micro vessels changes may result in diabetic retinopathy (retina damage), diabetic nephropathy (loss of kidney function) and diabetic neuropathy (nerve damage). Macro vessels changes may result in coronary heart disease, peripheral vascular disease and stroke (24, 25).

GDM may have consequences that affect not only mother but also the offspring (26, 27). The glucose molecule is small enough to cross over the placenta, whilst only a negligible amount of insulin can cross the placenta (28). High glucose levels stimulate fetal β cells to produce their own insulin that works as important growth hormone in fetal development. Excess glucose combined with hyperinsulinemia will result in increased storage of fat and proteins what may lead to macrosomia (29-31) (Figure 1). Macrosomia increases the risk of shoulder

dystocia and cesarean section. Right after cutting the navel cord, glucose levels drop drastically since the fetus no longer has direct access to the maternal blood and the high insulin levels can lead to dangerous low blood sugar (32, 33), which may result in fainting and severe brain injury. Intrauterine hyperglycemic environments may affect the offspring also in adult life, as the offspring have increased risk of developing obesity, T2DM and cardiovascular disease (34, 35).

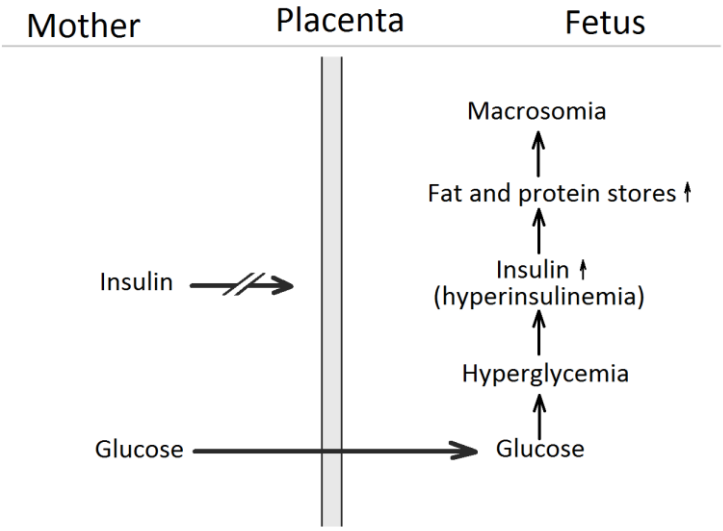


Figure 1: Schematic illustration of placental transport. Insulin cannot pass through the placenta like glucose molecules. High glucose levels in fetus blood result in high insulin levels, produced by fetus. Increased levels of glucose as fuel and insulin as anabolic hormone enhance the cellular uptake of glucose and result in storage of glucose as fats and proteins what in this way lead to increased birth weight. Inspired from Shakya et al. (29).

Link between environmental factors and diabetes

GDM and T2DM are recognized by increased hyperglycemia due to IR and impaired β cell function. The diseases share not only pathophysiology and genetic (36-38), but also risk factors such as obesity, weight gain, and increased age (12). Candidate gene studies have identified many genetic variants associated with T2DM and GDM (39), but they only explain a small part of the increased diseases risk. Hence, not only genetic factors, but also environmental factors, may play an important role. The understanding of epigenetics may help to explain how environmental factors may control gene expression by activation or silencing of some genes. Ling et al. explains in their review (40) how environment as metabolism, obesity, nutrition and aging is linked to the epigenetic alterations and chances to develop T2DM . It has also been suggested an association between a number of epigenetic modifications in different tissues and developing of GDM (14).

1.2 Epigenetics

Epigenetics is a collection of different mechanisms that can affect phenotype by regulation of gene expression without alterations of the DNA sequence (41, p.731). Epigenetic modifications can be inherited, but can also be lost or acquired during life (42, 43). Epigenetics play important role in cell differentiation: each cell in the body have the same DNA, which contains all genetic information, but only genes which code for proteins that are required for this specific cell are transcribed. Epigenetics regulate which genes will be transcribed and which will be silenced (44).

Chromatin

One important epigenetic tool to control gene expression is alterations of chromatin structure. DNA is a long molecule, which consists of two complementary polynucleotide chains that form a double helix structure. The DNA molecule is wrapped around octamer proteins that are called core histones and thus form smaller structure called nucleosomes (45, p.78-83, 199, 220-226). This structure of more compact DNA with associated proteins (both histones and non-histone proteins) forms chromatin (46, p.187). The density of DNA packaging decides the chromatin structure - more compact structures form heterochromatin and less dense form euchromatin. The difference in the structure plays a significant role in DNA transcription (47).

Heterochromatin does not allow binding of proteins, required for gene transcription to DNA, because there is no space when DNA is tightly packed (47). Nucleosome remodeling complex can bind to nucleosomes to remodel chromatin and get a more open DNA structure where transcription factors can bind (48).

Histones

The core of a nucleosome consists of four types of histone proteins: H2A, H2B, H3 and H4. One histone core comprises of one H3-H4 tetramer and two H2A-H2B dimers (49). The structure looks like a sphere with sticking out amino-terminal histone tails (Figure 2).

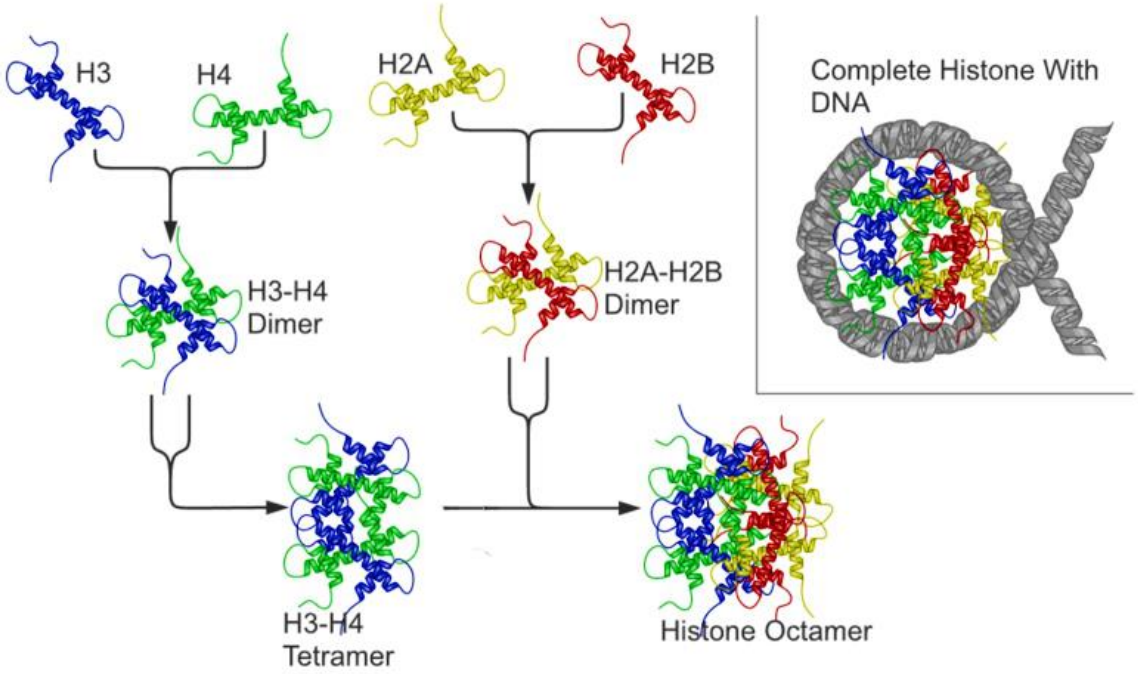


Figure 2: Core histone assembly into the nucleosomes.H3 and H4 proteins assemblies into one H3-H4 tetramer, but H2A and H2B proteins assemblies into two dimers and form one histone. Reprint with permission from (50).

The chromatin structure can be altered by different ways. The first one is by histone variant, proteins that can replace histone proteins, but may have another property. Such replacing and exchange play important role in different processes like proper embryonic development and germ cells development (51). Error in histone variants pattern can lead to cancer in human's cells (52).

Chromatin structure can also be altered epigenetically through histone modifications. Some amino residues in histone tails can be phosphorylated, acetylated, ubiquitinated or methylated. Thus, it has been observed for example, that phosphorylation and acetylation of histone tails are associated with gene activation, but deacetylation with gene inactivation. In this way, gene expression can be controlled by histone modifications (53).

DNA methylation

Another important epigenetic alteration to control gene expressions is DNA methylation. DNA structure can be covalently altered by adding methyl groups to the 5-carbon site of cytosine (C) bases. The process is catalyzed by DNA methyltransferase, where S-adenosyl-methionine (SAM) acts as the methyl donor (Figure 3) (54).

5-methylcytosine is mostly found in CG dinucleotides, often called CpG sites (5'-CpG-3', where "p" means phosphate bond between cytosine and guanine (G)). CpG sites tend to cluster and form regions with a high frequency of CpG sites, so called CpG islands (55).

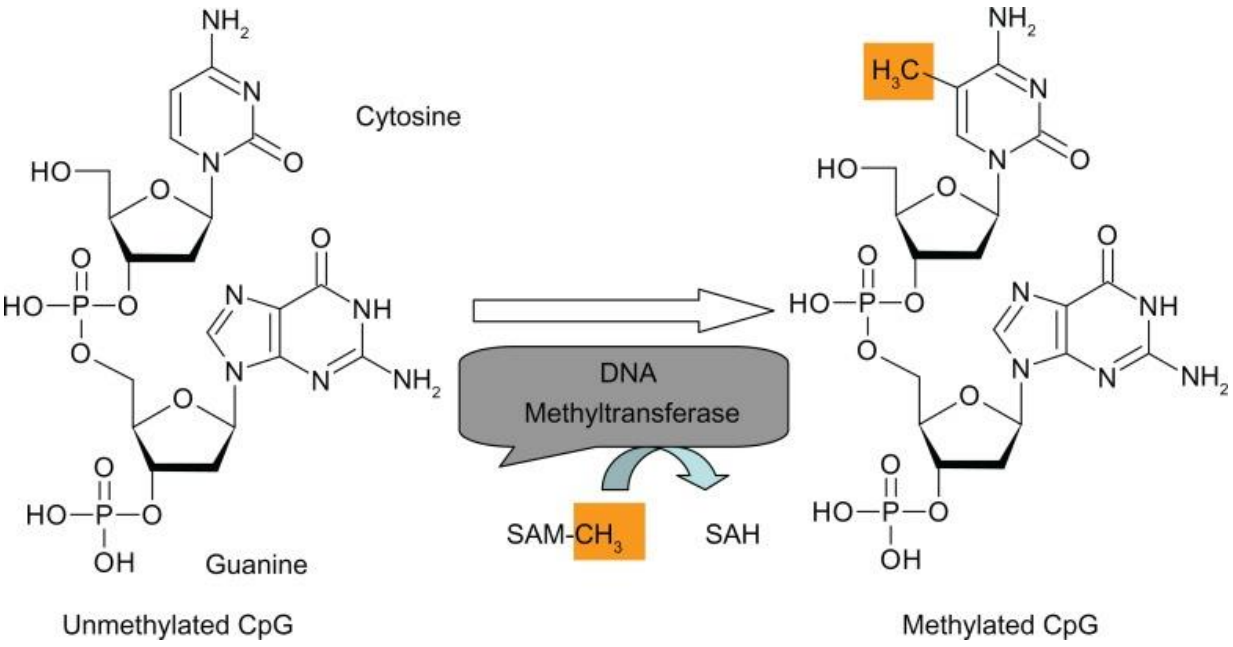


Figure 3: Methylation of the fifth position of cytosine in a CpG site catalyzed by DNA methyltransferase where the cytosine gets a methyl group from SAM donor. Reprint with permission from (56).

Since cytosine creates base pair with guanine, the CpG sites will be symmetrically reflected in both DNA strands and both cytosines can be methylated. This characteristic helps the methylation pattern be inherited further. During semiconservative DNA replication, the newly synthesized complementary DNA strand will not have DNA modifications like methylated cytosine (57). The symmetrical methylation can be kept by enzyme methyltransferase DNMT1. It recognizes methylated cytosine on one strand and methylate the symmetrical CpG site. In this way, the methylated pattern can be kept during cell division (46, p.405). The methylation pattern can be lost by reduced enzyme activity or by active demethylation when DNA demethylase removes methyl group from 5-methylcytosine.

During life, new methylations can be created in previously non-methylated CpG sites by *De novo* DNA methyltransferases DNMT3a and DNMT3b (58). There are four possible way to change or remain methylation pattern: CpG sites can be demethylated by demethylase, new methylation can occur by *De novo* DNA methyltransferase, the pattern can be kept after DNA replication by maintenance DNA methyltransferase or lost if the enzyme is not present (Figure 4).

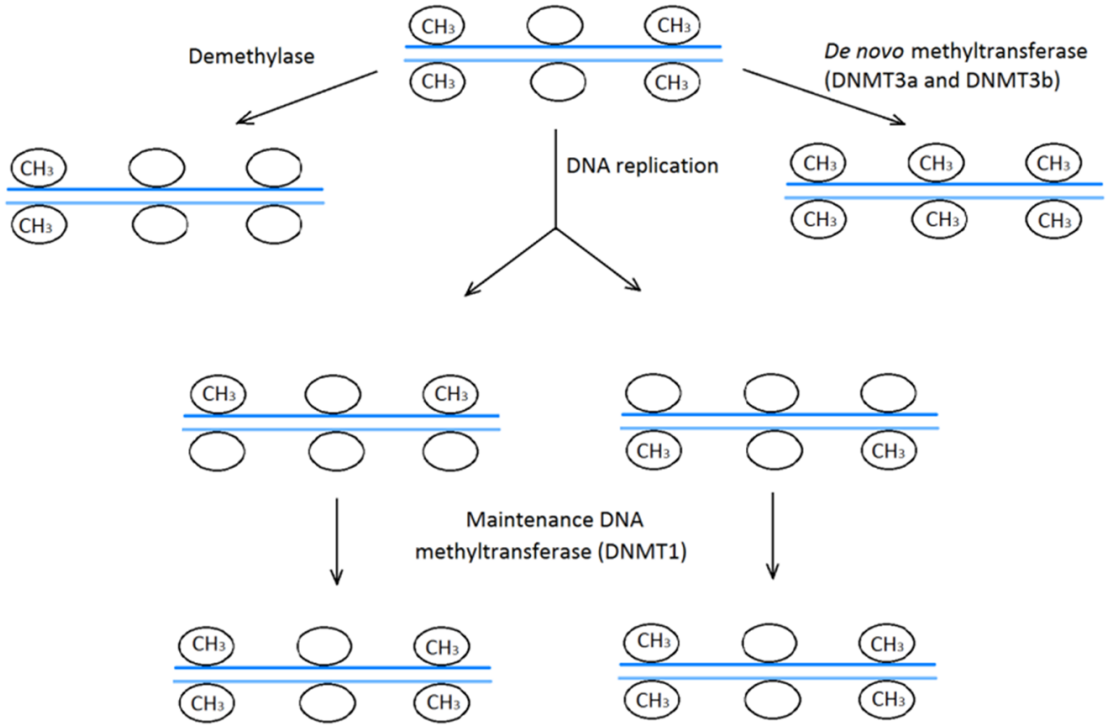


Figure 4: Summary of DNA methylation reactions. Rings on DNA with CH3 marks represent methylated CpG sites, empty rings on DNA represent CpG sites that potentially can be methylated. Inspired from Craig et al. (59, p.5).

DNA methylation often occurs in the promoter region of genes and is generally associated with decreased gene repression, whereas methylation in repressor region can increase gene expression (60). Methylation can also occur at a gene body and then lead to stimulation of gene transcription and play a role in alternative splicing (55).

DNA methylation represses gene expression by two mechanisms. First, methyl groups blocks protein-binding sites so that transcription factors cannot bind to start gene expression (60). Another way is by chromatin remodeling. DNA methylation in promoter region recruits certain proteins with methyl-CpG-binding domains (MBDs). MBDs recruit complexes that act as histone-modifying agents. The complexes contain histone deacetylases such as nucleosome remodeling deacetylase (NurD) complex and histone methyltransferases. Those enzymes alter chromatin stage to heterochromatin that leads to gene silencing (59, p.6, 61, 62).

The methylation pattern can be altered during life through different factors. Smoking is strongly associated with DNA methylation modifications (63, 64) and exposure for cigarette-

smoke of fetus can lead to a number of diseases in adult life (65). The pattern of DNA methylation can also be influenced by nutrition (66), age (67) and physical activity (68).

Epigenome-wide association study

Epigenome-wide association study (EWAS) is a genome-wide analysis of DNA methylations at a large number of CpG sites provided on Methylation BeadChip. The method is often used to explore associations between specific phenotypes, such as a trait or disease, and DNA methylation (69). According to a review by Moen et al. (14), a number of studies have focused on DNA methylation difference in newborns exposed to maternal hyperglycemia. The preferred tissues in those studies were placenta and umbilical cord blood. A small amount of studies has focused on variation of gene DNA methylation directly in pregnant women with GDM (70, 71). Those studies used maternal blood as tissue and found some differently methylated CpG sites, but the weakness of the studies was the small amount of participants. Our research group performed several EWAS as a part of the EPIPREG study (Role of epigenetics in diabetes development during pregnancy) that is using already collected data from a large population-based cohort study, the STORK Groruddalen study (Moen et al, unpublished). The DNA from white blood cells from 480 pregnant women was analyzed. The top two CpG sites was found to be different methylated associated with FPG (cg08098128 and cg14120215) and one CpG site associated with 2-h PG (cg19327414) (Moen et al. unpublished). The position of the CpG sites and related genes are presented in the Table 1. To ensure that those CpG sites and other methylation values obtained from Methylation BeadChip are not only artifacts, the Methylation BeadChip need to be technically validated with another method for site-specific CpG analysis. The widely used technic for this purpose is bisulfite pyrosequencing (BSP).

Table 1: CpG sites associated with FPG and 2-h PG found in the EPIPREG study (Moen et al, unpublished). The location information was found from the Genome Browser (72).

CpG site	Associated with	Location	Near/ inside gene	The gene is coding for protein
cg08098128	FPG	Chr. 2: 220 112 465	*STK16*	Serine/threonine kinase 16
cg14120215	FPG	Chr. 12: 113 841 668	*SDS*	Serine dehydratase
cg19327414	2-h PG	Chr. 6: 160 221 248	*PNLDC1*	Poly(A)-specific ribonuclease (PARN)-like domain containing 1

Other studies

A few other studies provided EWAS in association with plasma glucose in pregnancy or GDM, but with low number of participants (70, 71). However, some larger studies provided EWAS on white blood cells and reported some CpG sites associated with T2DM or diabetes relates traits.

The cross-sectional RODAM study (73) provided EWAS in 713 Ghanaian participants: 256 with T2D and 457 controls. The four CpG sites were found to be associated with T2DM: cg04816311, cg00574958, cg07988171 and cg19693031. Another study (74) provided EWAS among Indian Asian participants: 1074 with incident T2DM and 1590 controls. Seven CpG sites were found to be associated with T2DM: cg19693031, cg09152259, cg04999691, cg11024682, cg02650017, cg18181703 and cg06500161. The GOLDN study (75) provided EWAS among 837 non-diabetic European participants associated with T2DM related traits. The CpG site cg06500161 was found to be associated with insulin and HOMA-IR (Homeostatic Model Assessment of Insulin Resistance).

Purpose of further research

By exploring the association between epigenetic markers, hyperglycemia and GDM, it can be possible to better understand the underlying mechanisms leading to GDM and T2DM, predict and start treatment earlier to prevent severe complications, or improve treatment strategies.

2 Aims

The aims of the master project were:

- to technically validate the Illumina Methylation EPIC BeadChip (EPIC) using bisulfite pyrosequencing

- to explore how DNA methylation of top three candidate CpG sites from our own EWAS was associated with age, smoking status, GDM diagnosis, BMI, the homeostatic model assessment of insulin resistance (HOMA-IR) and β cell function (HOMA-B), insulin, total body fat mass and C-peptide values in the European and South Asian women participating in STORK Groruddalen

- to explore how DNA methylation of candidate CpG sites found in studies exploring T2DM were associated with FPG and 2-h PG in the EPIPREG sample

3 Methods

3.1 STORK Groruddalen and the EPIPREG study

Design

The STORK Groruddalen study is a population-based cohort study. The study included 823 healthy pregnant women living in Groruddalen who attended child health clinics in Stovner, Grorud and Bjerke administrative districts of Oslo, Norway, for antenatal care from May 2008 to May 2010 (76). Groruddalen is a multiethnic area where 48% in the population have minority background (77). Participating women attended three visits (Visit 1: gestational week <20, Visit 2: gestational week =28±2 and Visit 3: 10-14 weeks postpartum). Data was collected with questionnaires, anthropometric measurements, the oral glucose tolerance test (OGTT) and fasting blood samples collection. An overview over the measurements taken at the three visits is illustrated in the Figure 5.

Women were eligible to participate if they met the following conditions: 1) living in Stovner, Grorud or Bjerke city districts; 2) gave birth at one of two study hospitals (Akershus University Hospital or Oslo University Hospital, Ullevål); 3) gestation week at inclusion ≤ 20; 4) had healthy pregnancy without diseases requiring intensive hospital follow-up; 5) had the ability to give informed consent and 6) had the ability to communicate in Norwegian or any of the translated languages (Arabic, English, Somali, Sorani, Tamil, Turkish, Urdu or Vietnamese) (76).

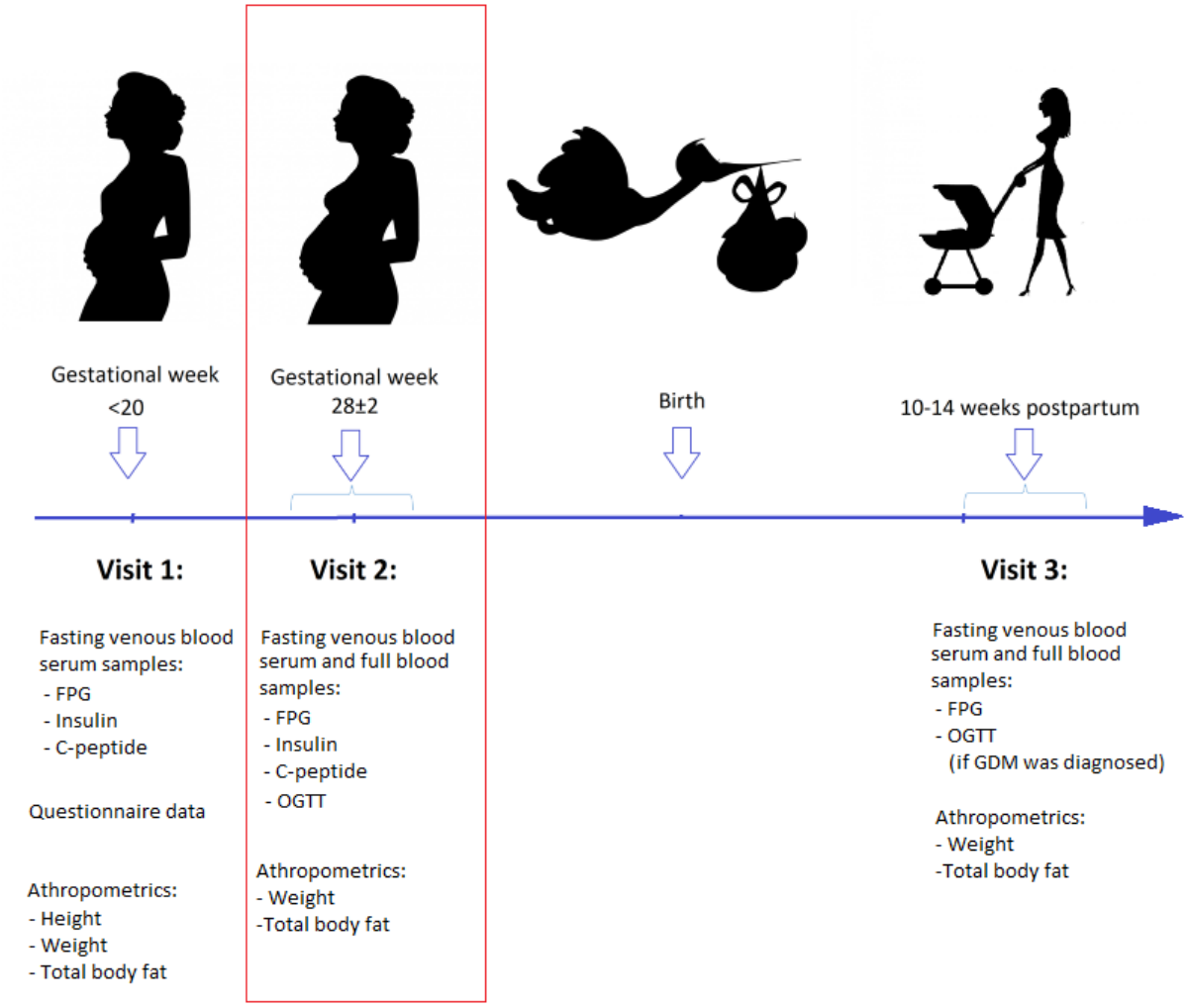


Figure 5: Overview of sample collection during the three visits In STORK Groruddalen study. Modified from publicdomainpictures.net (78-80). The red rectangle shows sample collection that was used in this master thesis plus the questionnaire data from visit 1.

The EPIPREG study quantified DNA methylation using the Infinium® MethylationEPIC BeadChip in Europeans (n=312) and South Asians (n=168) in the STORK Groruddalen study with decent DNA quality. This master thesis was a cross sectional study, using the EPIC data, the plasma glucose values and other measurements collected in gestation week 28±2.

Questionnaires and ethnicity

Age was calculated from mother's date of birth. The participant's marital status, education and smoking status are listed in Table 4. Gestational weeks were calculated using the first day of the last menstruation cycle. Information regarding family medical history included women with first-degree relatives with diabetes and questions about obstetric history including GDM

diagnosis in previously pregnancies were obtained from questionnaires at visit 1. Ethnicity was indicated based on genetic genome-wide analysis data. In this thesis, I present data from the two ethnic groups with DNA methylation data from the EPIPREG study: "European" (consists of women from Scandinavia, Eastern Europe, Western Europe and North America) and "South Asian" (women from Pakistan, Sri Lanka and India).

Anthropometric measurements

Mother's height was measured with a fixed stadiometer. Body weight and composition were measured with Tanita-weight BC 418 MA (Tanita, Tokyo, Japan) body fat analyzer (76). BMI was calculated from given height and body weight with formula: $\frac{body\ weight}{height^2}$ = (kg/m^2).

Blood measurements

After an overnight fast, venous blood samples with serum were collected during all visits. Fasting insulin and C-peptide were measured in serum by time-resolved fluorescence assays (DELFIA, PerkinElmer Life Sciences, Wallac Oy, Turku, Finland (81, 82)) at the Hormone Laboratory, Oslo University Hospital. The HOMA-IR and HOMA-B were calculated using C-peptide with the Oxford University HOMA Calculator 2.2 (83).

FPG was measured in serum after an overnight fast and 2-hour glucose in venous EDTA blood after OGTT on site using a patient-near method (HemoCue, Angelholm, Sweden). OGTT is used to determine how fast sugar leaves the blood after ingestion of a defined amount of glucose (84). After measuring the FPG, the participant drinks a solution of water with glucose and blood glucose is measured again after 2 hours (85). FPG values higher than normal will indicate impaired fasting glucose and 2-h PG higher than normal levels will indicate impaired glucose tolerance. Plasma glucose is measured in mmol/L in Norway. GDM was defined with the World Health Organization criteria proposed in 1999 (WHO1999) (86) with FPG ≥ 7.0 mmol/L and 2-h PG ≥ 7.9 mmol/L and with the International Association of Diabetes and Pregnancy Study Groups (IADPSG) criteria (new WHO criteria or WHO2013 criteria) (13) with FPG ≥ 5.1 mmol/L and 2-h PG ≥ 8.5 mmol/L (87).

DNA extraction

DNA extraction procedure was performed by Hormone Laboratory, Aker Hospital / Oslo University Hospital according to protocol (88). EDTA venous blood contains all blood cell types, but erythrocytes and thrombocytes lack nucleus, which result in DNA extraction from white blood cells only.

The blood cell underwent lysis after adding of Buffer A. After some tubes rotations, the samples were centrifuged. Supernatant was poured off and Buffer A was added again. Samples underwent a new centrifugation round. Supernatant was poured off. To provide denaturation of all proteins Buffer B and Buffer C were used. Samples were incubated overnight at 37°C. After incubation period, there was added saturated NaCl and samples were shaken vigorously for 15 seconds. Sodium ions attract to negatively charged DNA and thus temporary neutralize it (opportunity to be dissociated from the water). After centrifugation, the supernatant containing DNA was isolated. DNA was precipitated by adding ethanol. Precipitated DNA was contained with bent glass pipette and solved in water. (88)

DNA methylation analysis

The DNA methylation was quantified in white blood cells in 850,000 probe sites by project collaborators at Lund University Diabetes Centre in 2017, using the EPIC.

In short, the workflow consisted of following steps: (1) DNA bisulfite conversion, (2) DNA amplification, (3) enzymatic fragmentation, (4) DNA precipitation, (5) DNA resuspension, (6) hybridizing of DNA on BeadChip, (7) single base extension, (8) stain of BeadChip and (9) scan and image BeadChip.

Illumina BeadChip technology applied both Infinium I and Infinium II Methylation Assays (89). There was programmed in the Illumina which assay should be applied to specific CpG site.

The Infinium I Methylation Assay had two bead types designed for each CpG site to determine methylation state. Single stranded oligonucleotides were attached to the beads and were served as primers. The sequences of oligonucleotides was almost similar except the free ends (one with 3'-AC-5' dinucleotides on the end that corresponded and fit to unmethylated sample and one with 3'-GC-5' dinucleotides on the end, corresponding and fit to methylated

sample) (90). After hybridization, the primers that hybridized to DNA undergo single-base extension with labeled nucleotides and staining after that. The BeadChip was then ready to scanning (91) (Figure 6).

The Infinium II Methylation Assay had only one bead type with bounded primer per CpG site (89). Single base extension of primer with labeled nucleotides used bisulfite converted DNA as template what allowed to determine genotype for the sample. In case where CpG site was methylated, the G nucleotide incorporated with specific label and in case unmethylated locus the A nucleotide incorporated with own specific label. After staining the BeadChip was ready to scanning (92). Workflow and assays is illustrated in Figure 6.

GenomeStudio software was used to visualize information in form of graphs, scatter and line plots, and heat maps (89). The meffil R package was used to perform the following quality control and normalization (Moen et al, unpublished). As result, there was possible to determine methylation level of each CpG sites in the genome. A β value of the methylation level with range from 0 to 1 (from not methylated to completely methylated) was obtained for each women at each CpG site. To confirm these β values, the technical validation of the EPIC was performed in this thesis.

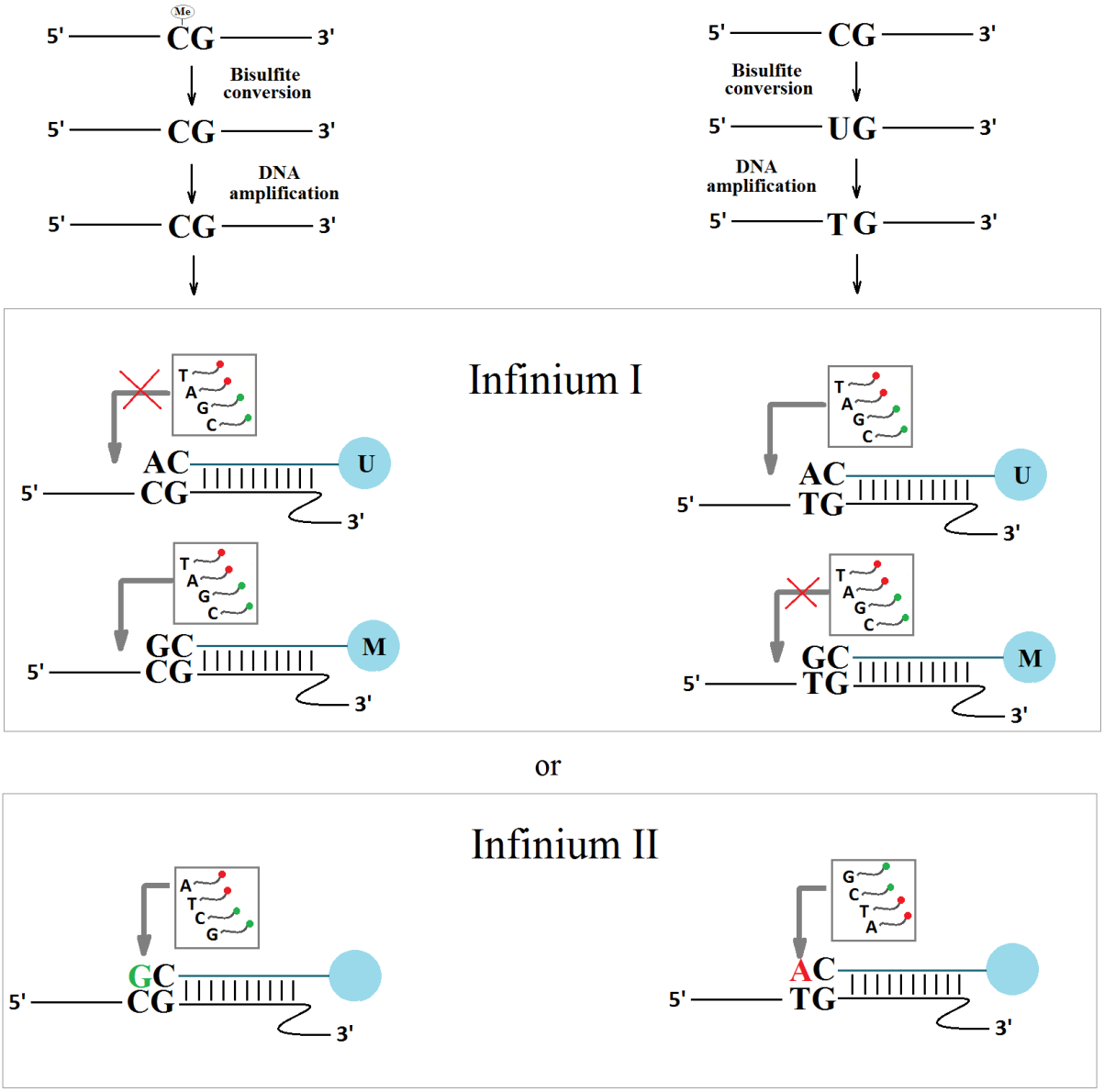


Figure 6: Preparation to and workflow of Infinium Methylation Assays. [1] Methylated sample locus with CG will stay remain CG after bisulfite conversion and DNA amplification. In Infinium I assay the sample will binds to primers bounded to two beads that represent methylated (M) and unmethylated (U) stages. The methylated sample will fit the primers with GC nucleotide on the free end that represent methylated state and allow single base extension with labeled nucleotides only on the primer what will be registered. In Infinium II assay only one bead type exist with primer what will be singe base extended with labeled nucleotides after binding of DNA sample. Since the locus is methylated, will labeled G incorporates what will be registered and indicate methylation. [2] Unmethylated sample locus will alter CG to UG after bisulfite conversion and to TG after DNA amplification. In Infinium I assay it will fit the primer with AC dinucleotide on the free end that correspond to unmethylated stage (U). In Infinium II assay the labeled A will incorporate, what will be registered and correspond unmethylated locus. Inspired from Illumina (89, 90, 93).

3.2 Validation study performed by the master student

Random samples of 30 women were selected for technical validation of the EPIC. Two CpG sites (cg14120215 and cg08098121) associated with FPG were selected to be validated by pyrosequencing. In addition, one CpG site associated with pre-pregnancy BMI (cg17148978) from the same EWAS results was chosen to be validated to another EPIPREG project. The BSP was performed by the master student at Akershus University Hospital.

The main principle of the BSP was sequencing-by-synthesis what was provided by detection and registration of each nucleotide that was included in newly synthesized DNA strand. After nucleotide incorporation, the pyrophosphate (PPi) was released which led to visible light generation and determination of DNA sequence (94).

DNA samples were first bisulfite converted according to Qiagen Bisulfite conversion protocol (95). Short DNA sequences that contained CpG site of interest was amplified by polymerase chain reaction (PCR) and pyrosequensed. In this way, each CpG site was validated in all 30 DNA samples. In addition, it has been performed Whole Bisulfitome procedure in all DNA samples where they were amplified with possibility to store them longer.

DNA bisulfite conversion

The DNA of interest was first treated with bisulfite before the pyrosequencing. Bisulfite turns cytosines into uracils (U) (C→U) in multiple steps. Methyl groups in methylated cytosines protected them from conversion and they remained unchanged (Figure 8).

To perform bisulfite reaction, 500 ng of each DNA sample was added in 200 µl tubes together with bisulfite solution and DNA protect buffer. The volume was adjusted with RNAse free water. At right conditions, after mixing the solution will change the color from green to blue. In this case, there was performed bisulfite conversion in thermal cycler with condition from the Table 2.

Table 2: Thermal cycler conditions under DNA bisulfite conversion.

Step	Time	Temperature
Denaturation	5 min	95°C
Incubation	15 min	60°C
Denaturation	5 min	95°C
Incubation	15 min	60°C

The converted DNA samples were cleaned-up in a multiple step procedure. The BL buffer with carrier RNA and absolute ethanol were added to each converted DNA sample in Eppendorf tube. After vortexing and spinning, the solutions were transferred into MinElute DNA spin columns placed in tubes. After 1 min centrifuge the flow-through was thrown and wash buffer was added into columns. The tubes with the columns were centrifuged at 1 min, the flow-through was thrown and desulfonation buffer was added. The samples were incubated in 15 min at room temperature and centrifuged for 1 min. The flow-through was thrown and wash buffer was added. After 1 min centrifuge, the flow-through was thrown and last step was repeated with wash buffer. The absolute ethanol was added and tubes were centrifuged for 1 min, the flow-through was thrown. After the columns were dried, they were placed in new Eppendorf tubes and EB buffer was added. After 1 min incubation at room temperature, the tubes with the columns were centrifuged. The flow-through was eluated one more time thought spin columns by centrifugation to increase DNA concentration. Eluated DNA was stored at -20°C. The workflow of the procedure presented in Appendix III.

Whole bisulfitome

The procedure was performed according to EpiTect® Whole Bisulfitome Handbook (96). The nuclease free water was added to 150 ng of bisulfit converted DNA to adjust the volume up to 10 µl. There was added 30 µl of EpiTect Amplification Master Mix (consisting of 29 µl of EpiTect WBA Reaction Buffer and 1 µl of REPLI-g Midi DNA Polymerase) to the each sample. The solutions were incubated at 28°C for 8 hours. After the incubation, the REPLI-g Midi DNA Polymerase was inactivated by heating the samples up to 95 °C for 5 min.

The new concentration of each probe was measured with Quant-iT™ double-stranded DNA (dsDNA) Assay Kit according to protocol (97). Into each microplate well there was eluted 200 µl of freshly made working solution. In some wells was also eluted 10 µl of λ DNA

standards, and 1 μl of DNA samples in doublets in others. Using microplate reader, it has been determined DNA amount in the each well. Amplified bisulfit converted DNA was stored at -20°C. Some of those samples were used in PCR and further pyrosequencing when we have run out original DNA samples.

PCR

The next step was PCR amplification. DNA sequences in both DNA strands was changed, and the PCR was therefore performed on top and bottom stands independently. Two primers were used to separate the strands, where one of them was biotinylated (Figure 8).

All CpG sites of interest lie in different positions in the genome. Therefore, there was necessary to amplify the two DNA sequence that contains the CpG sites independently. In this way, all 30 samples were PCR amplified for each CpG site.

PCR Master Mix was made for CpG site containing: PyroMark PCR MM, Coralload concentrate, forward and reverse PCR primers for specific CpG site and RNAse free water. The concentration of each reagent is presented in Appendix IX.

The PCR Master Mix was added to the each DNA sample (with total volume 25 μl), the solution was mixed and the 45 PCR cycles were performed on thermal cycler with condition from Table 3.

Table 3: Thermal cycler conditions under PCR.

Components	Time	Temperature
Initial PCR activation step	15 min	95°C
3 step cycle:		
Denaturation	30 sec	94°C
Annealing	30 sec	56°C
Extension	30 sec	72°C
Number of cycles	45	
Final extension	10 min	72°C

PCR was performed in 30 samples in doublets (60 probes) with two positives controls: Unmethylated converted DNA, Methylated converted DNA and a negative control. There was used 20 - 40 ng input DNA (20 ng for cg17148978 and cg14120215, and 40 ng for cg08098128) and specific PCR primer for each CpG site sample. In cases with bad results

from pyrosequencing, the PCR step was repeated with same conditions or with increased DNA input.

The amplified DNA was tested by the gel electrophorese in TAE buffer at 80 V for 30-40 min in 1.5% agarose (example in the Figure 7, reagents in Appendix X).

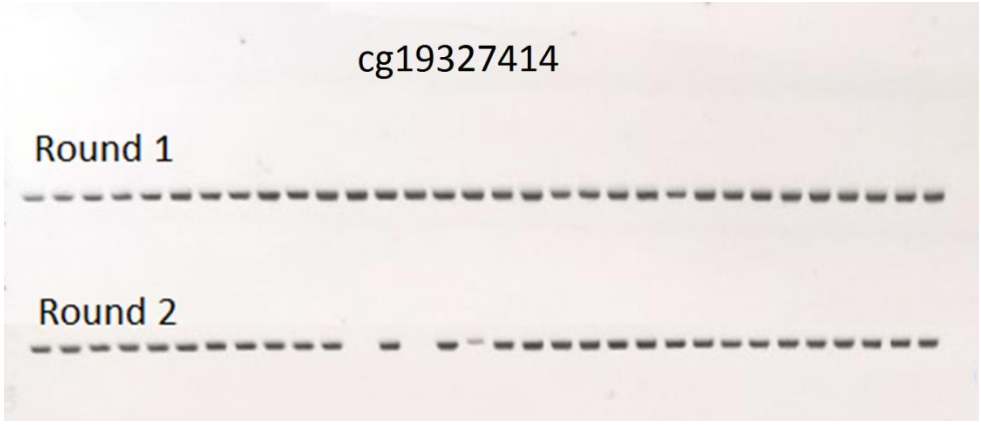


Figure 7: Results from agarose gel after PCR amplification of 1-30 DNA samples, unmethylated converted control, methylated converted control and negative control from cg14120215. The DNA samples was amplified and analyzed in doublets (30 DNA samples in Round 1 and 30 DNA samples in Round 2). DNA samples 12 and 14 from Round 2 lack in the picture. The PCR for those samples was repeated.

Pyrosequencing

PyroMark Q48 Autoprep was used in the procedure according to PyroMark Q48 Autoprep User Manual (98). The run file for the each run was created using PyroMark Q48 Autoprep Software, downloaded into the USB stick and inserted into the instrument. The volume needed of each reagents for the run was showed on the display. The room temperature reagents was loaded in the instrument cartridges: all four nucleotides (A, C, G, T) in the different cartridges, enzyme, substrate, sequencing primer and buffers. The DNA samples with magnetic beads was loaded into the PyroMark disk. After the disk was placed in the instrument, the pyrosequencing run could be started.

Primarily, the double stranded PCR amplicons were converted to single strand DNA fragments. Due to PyroMark Binding Buffer, the strands with biotin were bound to streptavidin-coated beads that held the strand on vacuum filtration tips. The complimentary strands were separated by PyroMark Denaturation Solution, while PyroMark Wash Buffer allowed the removing of strands without biotin (99). The biotinylated strand served as template to further pyrosequencing (94). (Figure 8)

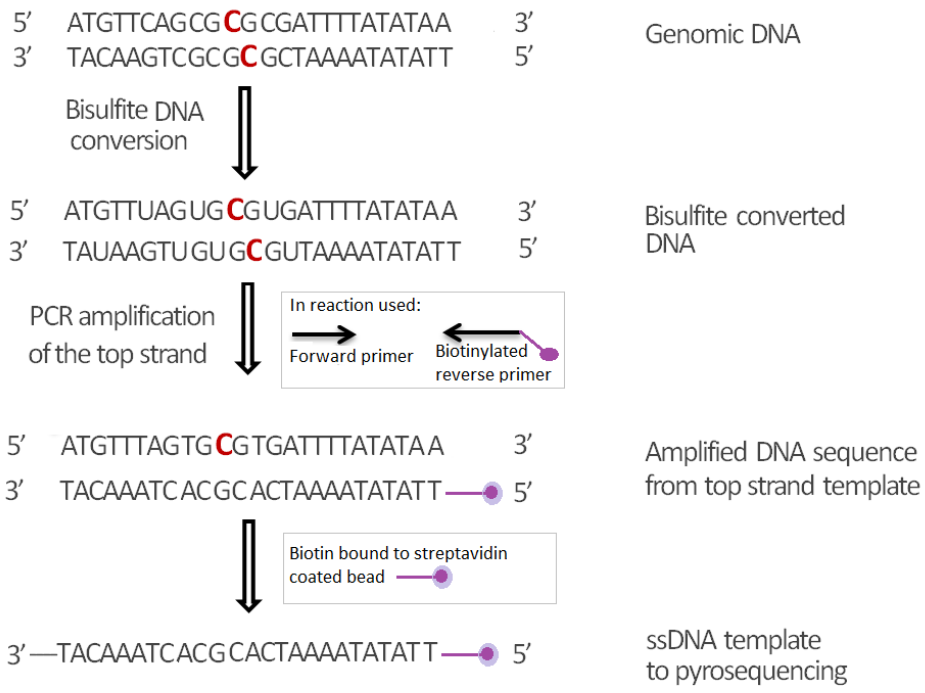


Figure 8: Preparation of single-stranded DNA (ssDNA) template to pyrosequencing. Red "C" in DNA stand represents 5-methylcytosine. All unmethylated cytosines are converted to uracils after the bisulfite treatment and the top and bottom DNA strands have not longer complete complementary sequences. The two strands therefore amplified by PCR independently. For simplicity, only the top strand amplification is shown here. Bisulfite converted DNA is PCR amplified with a primer pairs where one of them (in the example reverse primer labeled with purple sphere) is biotinylated. Biotin provides binding of amplified DNA to streptavidin-coated beads. The process is needed to convert the amplified dsDNA fragments into ssDNA fragments ready to pyrosequencing. Inspired from Craig et al. (59, p.298, 299, 303, 304).

After the single strand template generation, the four solutions (one with adenine (A) nucleotides, one with thymine (T), one with cytosine and one with guanine) were added to solution with single-stranded DNA (ssDNA) template one at the time. If a nucleotide was not incorporated in the new strand there was washed out (Figure 9A) and solution with another nucleotide was added. If the new nucleotide fit the growing strand (Figure 9B), DNA polymerase included it in the strand and pyrophosphate was released. This process further initiated a chain of reactions that led to a light release. The intensity was be proportional to the amount of each incorporated nucleotide type (100).

The first reaction was provided by enzyme ATP sulfurylase which converted pyrophosphate in ATP in absence of adenosine 5′-phosphosulfate (APS) (Figure 9C) (101). The released ATP served as energy source to convert luciferin to oxyluciferin by enzyme luciferase. Depending on amount of ATP from the previous reaction (the same amount as PPi) a given intensity of visible light was increased (Figure 9D) (94). The light was registered and results

were showed in a graph called pyrogram (Figure 9E). Length of peaks was proportional to the amount of released light. “A” in the pyrogram will present adenines in DNA sequence before the bisulfite treatment, “G” will present guanines, “C” will present methylated cytosines and “T” will present thymines or unmethylated cytosines (because of bisulfite conversion).

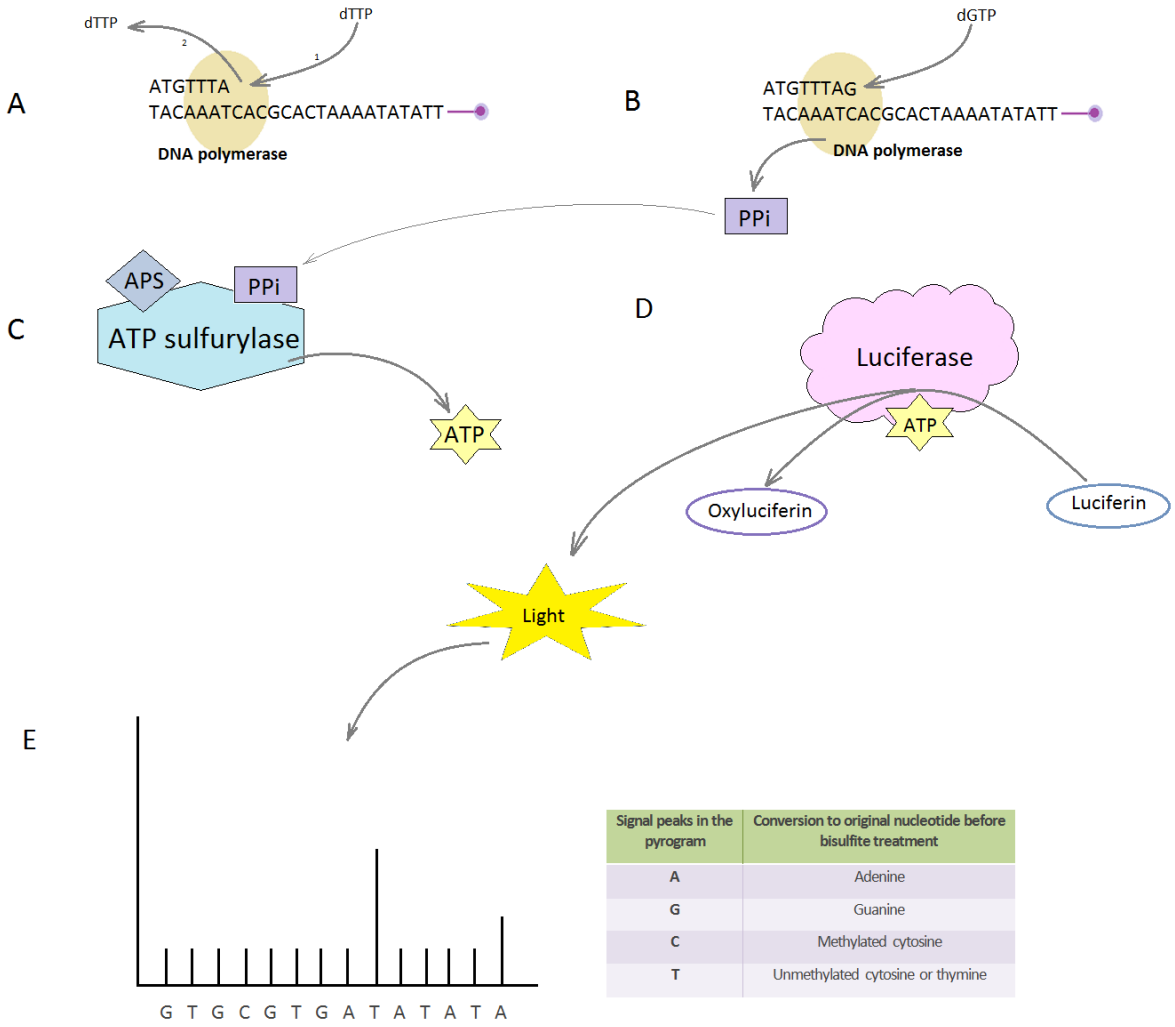


Figure 9: Schematic presentation of pyrosequencing. (A) The sequence primer is annealed; DNA polymerase is ready to drive synthesis. First is dTTP nucleotide added, but it does not match to cytosine on the template strand, so dTTP is degraded and washed out. (B) The new nucleotide dGTP is added. It matches to template and therefore incorporates. The pyrophosphate releases. (C) ATP sulfurylase convert PPi into ATP in absence of APS. (D) ATP serves as fuel to oxidation of luciferin intro oxyluciferin by enzyme luciferase. During this reaction the visible light is released. (E) The registration of light intensity make possible to determine nucleotide sequence and show it in pyrogram. Inspired from Craig et al. and Bassil et al. (59, p.304, 94).

After the pyrosequencing was finished, the data from the run was transferred into USB stick. The new run file was analyzed using PyroMark Q48 Autoprep Software (102). Methylation level of each CpG site was displayed in the pyrogram in percentage. The quality assessment was visualized as a color bar with multiple variants: (1) Blue – passes (good quality), (2)

Yellow – check (medium quality, must be repeated), (3) Red – failed (bad quality, must be repeated) (Figures 10 and 11).

Pyrosequencing troubleshooting

All assays for each CpG site underwent troubleshooting under pyrosequencing due to poor quality (yellow and red in the color bar in the Pyrogram) (Figure 10). The PyroMark instrument underwent two times “Injector test” and one “Pyrophosphatase clean” to make sure that the instrument’s work is correct. The concentration of sequencing primers was gradually increased with further test runs. The final concentration of sequencing primers was increased from 4 µM (Qiagen’s prescription) up to 6 µM for cg17148978, 10 µM for cg14120215 and 8 µM for cg08098128.

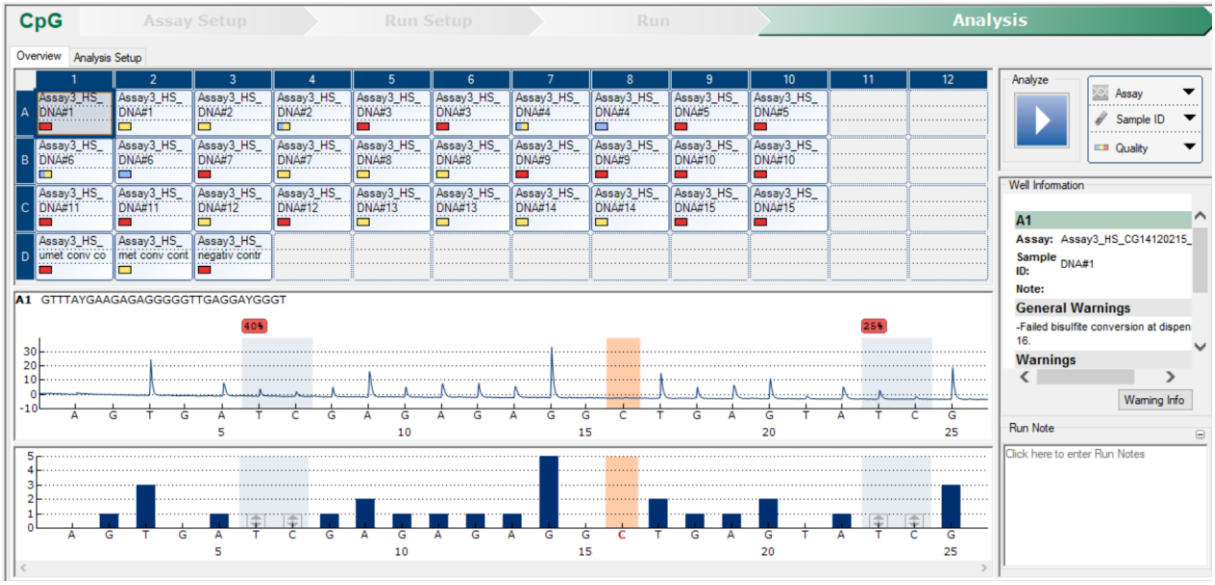


Figure 10: Results from one pyrosequencing run visualized with PyroMark Q48 Autoprep Software. The cg14120215 was analyzed with 20 ng DNA input in PCR. The concentration of sequencing primer was 4 µM. The pyrogram in the middle panel and the dispensation order display in the down panel represent highlighted sample DNA#1. The CpG site of interest is shown in the second position (order position 23-24). Most of the samples show poor quality with yellow and red identifying color and different warning massages that were taken into account to next pyrosequencing run.

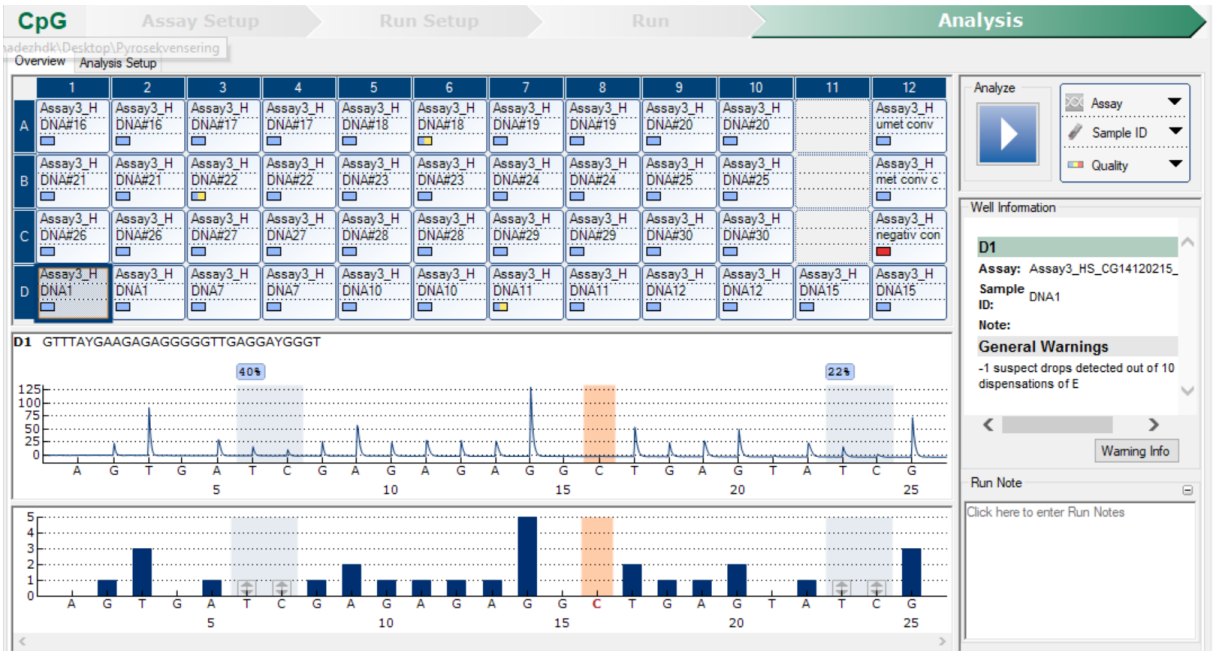


Figure 11: Results from another pyrosequencing run repeated on cg14120215 with 6 µM sequencing primer. Most of the samples show good results with blue identifying color. Some samples with yellow color (DNA#11, DNA#18 and DNA #22) was analyzed again until good results.

3.3 Statistics

The statistical analysis was performed by the student using R 3.4.4. The package TableOne (https://cran.r-project.org) was used to explore the association between phenotypes across tertiles of DNA methylation. The chi-squared test were applied for categorical variables, ANOVA (analysis of variance) test for continuous variables and Kruskal-Wallis test for non-normal variables. R codes are attached in the Appendix II "TableOne".

To define the reproducibility of the BSP method, we used the methylation values from all runs. It has been defined the difference between the minimal and maximal values of each DNA sample and the average difference was calculated for each CpG site in Excel.

The Pearson's correlation analysis, to compare EPIC with BSP was done in R. The average values for each DNA sample after all BSP runs were compared with β values get from the EPIC. R codes are attached in the Appendix II "EPIC – BSP correlation analysis". Since the β values range from 0 to 1, while BSP-values represent a percentage ratio, the β values was multiplied by 100 to convert them to percentages as BSP-values. The difference between methylation values from both methods for each DNA sample were calculated and the overage

difference was found for each CpG site in Excel. The difference between EPIC and BSP-values was presented as absolute values.

DNA methylation at candidate CpG sites found in EPIPREG study were sorted in tertiles according to low, medium and high methylation level at each CpG site for Europeans and South Asians and presented in the Tables 5 – 7. Due to similar pathophysiology of T2DM and GDM, CpG sites from previously published EWAS were also analyzed in STORK Groruddalen cohort (Tables 8 - 9). The variables as age, smoking, GDM, FPG, 2-h PG, BMI, HOMA-IR, HOMA-B, C-peptide, fat mass and insulin were used in EPIPREG sample. Only the FPG and 2-h PG values from 28±2 gestational week were sorted in tertiles for CpG sites found in other studies. In addition, the insulin and HOMA-IR-values from 28±2 gestational week were sorted and analyzed by cg06500161 methylation data (Table 10).

The categorical variables (smoking and GDM diagnosis) are presented as mean ± standard deviation (SD) or n (%), where the chi-squared test were applied. The continuous variable (age, FPG, 2-h PG, BMI, HOMA-B, fat mass and C-peptide) are presented as mean ± SD, where the ANOVA test was applied. The non-normal variables (insulin and HOMA-IR) are presented as median, where the Kruskal-Wallis test was applied. A p-value of 0.05 was used as significant. R codes are attached in the Appendix II "Association analysis".

3.4 Ethics

All statistical analyses were performed in Tjenester for Sensitive Data (TSD). The project has all necessary approvals to perform the planned analyses (REK 478-07247a 1.2007.894, The Norwegian Supervisory Authority and the Norwegian Institute of Public Health (biobank)).

4 Results

4.1 Cohort characteristics

The numbers of participants in this project was 314 Europeans (84.9 % of European participants attending the STORK Groruddalen study) and 166 South Asians (83 % of the South Asian participants). The results from measurements, questionnaires and blood analysis are presented in Table 4.

Europeans in the cohort were older and had lower parity than South Asian. Europeans also had higher education level and higher number of smokers (prior to pregnancy or current). The number of women diagnosed GDM were higher in South Asians, both with WHO1999 and WHO2013 criteria. They had also higher degree of diabetes in family.

The blood analysis in gestational week 28±2 indicated higher-values of FPG, 2-h PG, HOMA-IR, HOMA-B, C-peptide and insulin in South Asian women, while BMI and total body fat mas were higher in Europeans.

Table 4: Characteristics of the STORK Groruddalen cohort stratified by ethnic groups.

	Europe n = 312	South Asia n = 168	p-value
Age (years)	30.10 ± 4.62	28.17 ± 4.62	<0.001
Nulliparous (%)	161 (51.6)	68 (40.5)	0.026
Marital status (%)			<0.001
In relationship/ married	297 (95.2)	166 (98.8)	
Single	15 (4.8)	2 (1.2)	
Education level (%)			<0.001
<9 years schooling	14 (4.5)	32 (19.2)	
1-3 years secondary school	101 (32.5)	83 (49.7)	
University/college ≥ 4 years	196 (63)	52 (31.2)	
Smoking (%)			<0.001
Current	20 (6.5)	1 (0.6)	
Smoked 3 months prior to pregnancy	80 (25.9)	2 (1.2)	
Former	87 (28.2)	9 (5.4)	
Never	122 (39.5)	155 (92.8)	
Week at inclusion	14.21 ± 2.39	15.63 ± 3.87	<0.001
GDM with WHO1999 (%)	37 (11.9)	25 (15.1)	0.396
GDM with WHO2013 (%)	76 (24.4)	70 (41.7)	<0.001
Previously GDM (%)	3 (0.96)	8 (4.76)	0.020
A first-degree relative with diabetes, n (%)	43 (14.0)	79 (47.3)	<0.001
Fasting glucose, (mmol/L)	4.36 ± 0.50	4.49 ± 0.48	0.008
2-hour glucose, (mmol/L)	5.57 ± 1.43	5.98 ± 1.63	0.006
BMI, (kg/m2)	27.75 ± 4.72	26.83 ± 4.07	0.033
HOMA-IR	1.48 [1.16, 1.90]	1.76 [1.41, 2.31]	<0.001
HOMA-B	180.59 ± 51.12	186.66 ± 53.92	0.227
C-peptide, (pmol/L)	786.42 ± 345.77	905.05 ± 357.73	<0.001
Insulin, (pmol/L)	48.00 [33.00, 70.25]	71.00 [57.00, 100.50]	<0.001
Total body fat mass, (kg)	29.62 (9.66)	26.24 (8.23)	<0.001

The categorical variables (nulliparous, marital status, education level, smoking, GDM diagnosis and family diabetes) are presented as mean ± standard deviation (SD) or n (%), where the chi-squared test were applied. The continuous variable (age, inclusion week, FPG, 2-h PG, BMI, HOMA-IR, HOMA-B, fat mass and C-peptide) are presented as mean ± SD, where the ANOVA test was applied. The non-normal variables (insulin and HOMA-IR) are presented as median, where the Kruskal-Wallis test was applied.

4.2 **Technical validation**

Figure 12 shows the comparison of DNA methylation quantified with the EPIC versus BSP from identical CpG sites. The methylation variation seems to be low for cg08098128 and cg17148978 and higher for cg14120215 (Appendix I, Tables 11 – 13). The correlation between DNA methylation with BSP and EPIC was high (r = 0.64, p = <0.001).

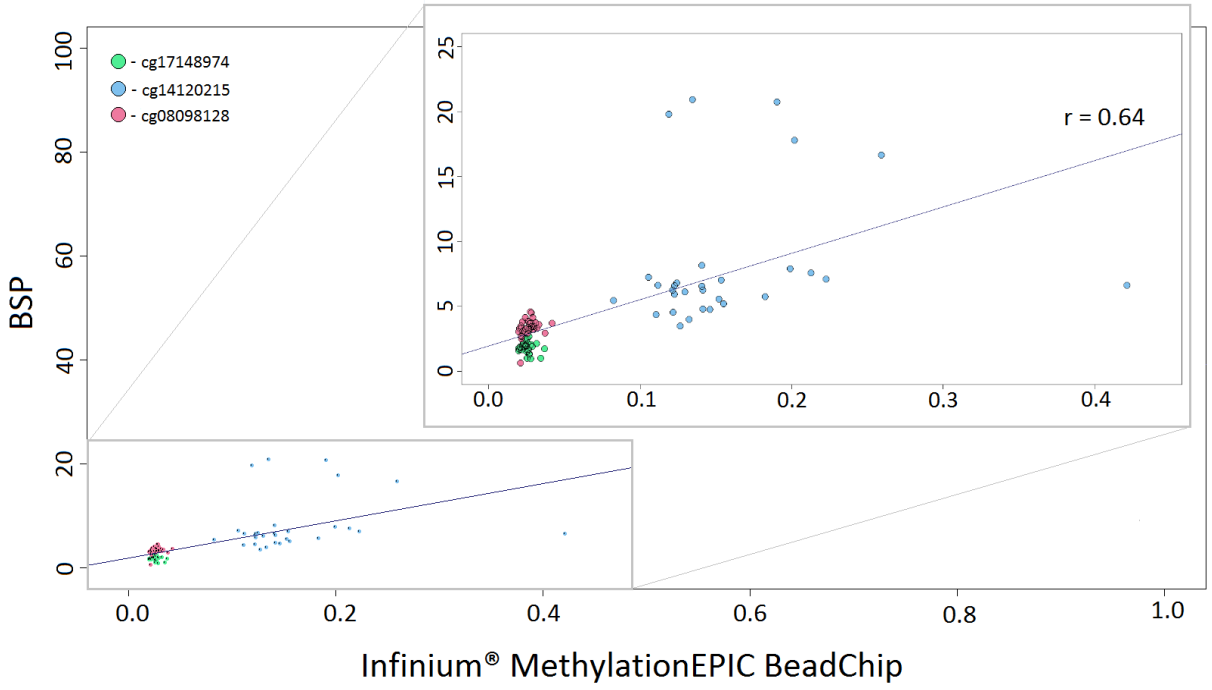


Figure 12: Scatter plot showing the relationship between DNA methylation assesses with the EPIC on x-axis and with BSP with percentage DNA methylation on y-axis for 30 DNA samples for cg17148978, cg14120215 and cg08098128 (points are labeled with different colors). r = 0.64, p < 0.001.

4.3 DNA methylation analysis

4.3.1 Candidate CpG sites from EPIPREG

BMI at gestational week 28±2 seems to be increased with increased methylation of cg08098128 in Europeans (p-value 0.046, Table 5). There was not found any significant results for South Asians.

In Europeans, the age was decreased with increased methylation level of cg14120215 (Table 6). There was not found any significant results for South Asians.

There was not found any significant results associated with cg19327414 methylation (Table 7), besides 2-h PG in Europeans with significant p-value of 0.044, but the CpG is associated with 2-h PG and the result was.

Table 5: Association between tertiles of DNA methylation of cg08098128 associated with FPG and other related factors in Europeans and South Asians.

Europeans	**[0.0158 - 0.024]**	**[0.024 - 0.0295]**	**[0.0295 - 0.174]**	**P**
n	103	102	102	
Age	29.74 (4.00)	30.47 (5.01)	30.10 (4.71)	0.522
Smoking, never (%)	39 (38.2)	44 (43.1)	35 (35.0)	0.113
GDM with WHO 1999 (%)	8 (7.8)	17 (16.7)	11 (10.8)	0.132
GDM with WHO 2013 (%)	23 (22.3)	23 (22.5)	27 (26.5)	0.736
Fasting glucose, (mmol/L)	4.30 (0.40)	4.35 (0.51)	4.42 (0.55)	0.187
2-hour glucose, (mmol/L)	5.64 (1.40)	5.64 (1.59)	5.43 (1.29)	0.515
Pregnancy BMI, (kg/m2)	26.89 (3.68)	27.67 (4.43)	28.53 (5.77)	0.046
HOMA-IR	1.40 [1.13, 1.85]	1.48 [1.18, 1.99]	1.50 [1.17, 1.85]	0.720
HOMA-B	183.26 (53.37)	181.67 (48.96)	176.68 (50.65)	0.632
Insulin, (pmol/L)	49.00 [33.00, 71.00]	46.00 [33.75, 67.25]	48.00 [34.25, 70.75]	0.845
C-peptide, (pmol/L)	787.78 (394.43)	770.70 (293.42)	786.67 (334.68)	0.925
Total body fat mass, (kg)	28.18 (7.34)	29.27 (9.39)	31.08 (11.75)	0.101
South Asians	**[0.0175 - 0.0265]**	**[0.0265 - 0.0313]**	**[0.0313 - 0.285]**	
n	55	55	55	
Age	27.91 (4.52)	28.40 (4.53)	28.33 (4.78)	0.834
Smoking, never (%)	49 (90.7)	51 (92.7)	52 (94.5)	0.712
GDM with WHO 1999 (%)	10 (18.5)	4 (7.3)	10 (18.5)	0.160
GDM with WHO 2013 (%)	24 (43.6)	23 (41.8)	21 (38.2)	0.839
Fasting glucose, (mmol/L)	4.52 (0.51)	4.49 (0.42)	4.46 (0.50)	0.828
2-hour glucose, (mmol/L)	6.00 (1.69)	5.95 (1.36)	5.96 (1.88)	0.986
Pregnancy BMI, (kg/m2)	27.46 (4.73)	26.64 (3.81)	26.48 (3.66)	0.411
HOMA-IR	1.89 [1.48, 2.28]	1.70 [1.34, 2.32]	1.75 [1.49, 2.23]	0.811
HOMA-B	184.41 (42.45)	182.30 (42.41)	193.98 (72.99)	0.495
Insulin, (pmol/L)	75.00 [57.50, 103.50]	66.00 [49.00, 100.00]	71.00 [58.00, 94.00]	0.728
C-peptide, (pmol/L)	902.55 (324.96)	884.55 (347.70)	932.25 (409.86)	0.788
Total body fat mass, (kg)	27.46 (10.22)	25.28 (7.37)	26.06 (6.78)	0.395

The categorical variables are presented as mean ± standard deviation (SD) or n (%), where the chi-squared test were applied. The continuous variables are presented as mean ± SD, where the ANOVA test was applied. The non-normal variables are presented as median, where the Kruskal-Wallis test was applied.

Table 6: Association between tertiles of DNA methylation of cg14120215 associated with FPG and other related factors in Europeans and South Asians.

Europeans	**[0.0636 - 0.131]**	**[0.131 - 0.16]**	**[0.16 - 0.458]**	**p**
n	103	102	102	
Age	31.68 (3.63)	29.97 (4.86)	28.64 (4.70)	<0.001
Smoking, never (%)	41 (39.8)	42 (42.0)	35 (34.7)	0.079
GDM with WHO 1999 (%)	14 (13.6)	8 (7.8)	14 (13.7)	0.329
GDM with WHO 2013 (%)	23 (22.3)	19 (18.6)	31 (30.4)	0.130
Fasting glucose, (mmol/L)	4.37 (0.42)	4.32 (0.50)	4.38 (0.55)	0.599
2-hour glucose, (mmol/L)	6.00 (1.47)	5.13 (1.19)	5.60 (1.50)	<0.001
Pregnancy BMI, (kg/m2)	28.41 (4.63)	26.99 (4.35)	27.67 (5.14)	0.102
HOMA-IR	1.48 [1.19, 1.90]	1.40 [1.14, 1.81]	1.48 [1.16, 1.98]	0.607
HOMA-B	179.15 (42.44)	181.88 (61.42)	180.55 (47.58)	0.930
Insulin, (pmol/L)	48.50 [34.00, 67.00]	46.00 [30.00, 68.00]	48.00 [36.00, 76.50]	0.431
C-peptide, (pmol/L)	776.69 (307.01)	761.92 (343.73)	807.00 (375.67)	0.638
Total body fat mass, (kg)	31.09 (9.75)	28.69 (9.14)	28.70 (10.04)	0.128
South Asians	**[0.0693 - 0.132]**	[**0.132 - 0.164]**	[**0.164 - 0.398]**	
n	55	55	55	
Age	28.53 (3.93)	28.64 (4.97)	27.47 (4.79	0.342
Smoking, never (%)	48 (88.9)	53 (96.4)	51 (92.7)	0.328
GDM with WHO 1999 (%)	5 (9.3)	10 (18.2)	9 (16.7)	0.373
GDM with WHO 2013 (%)	20 (36.4)	26 (47.3)	22 (40.0)	0.496
Fasting glucose, (mmol/L)	4.42 (0.40)	4.51 (0.50)	4.54 (0.52)	0.403
2-hour glucose, (mmol/L)	5.63 (1.41)	6.06 (1.72)	6.25 (1.74)	0.134
Pregnancy BMI, (kg/m2)	26.97 (4.26)	27.40 (4.23)	26.21 (3.75)	0.301
HOMA-IR	1.71 [1.25, 2.14]	1.82 [1.48, 2.47]	1.82 [1.48, 2.27	0.386
HOMA-B	183.81 (46.21)	186.96 (41.66)	189.71 (71.03)	0.852
Insulin, (pmol/L)	66.00 [41.00, 92.50]	80.00 [61.25, 106.00]	68.50 [52.75, 92.25]	0.164
C-peptide, (pmol/L)	852.60 (334.13)	945.50 (364.97)	921.28 (380.81)	0.379
Total body fat mass, (kg)	26.33 (9.11)	27.21 (8.47)	25.28 (7.21)	0.496

The categorical variables are presented as mean ± standard deviation (SD) or n (%), where the chi-squared test were applied. The continuous variables are presented as mean ± SD, where the ANOVA test was applied. The non-normal variables are presented as median, where the Kruskal-Wallis test was applied.

Table 7: Association between tertiles of DNA methylation of cg19327414 associated with 2-h PG and other related factors in Europeans and South Asians.

Europeans	**[0.721 - 0.927]**	**(0.927 - 0.936]**	**(0.936 - 0.952]**	**p**
n	103	102	102	
Age	30.66 (4.21)	29.69 (5.01)	29.95 (4.50)	0.291
Smoking, never (%)	41 (40.2)	36 (35.6)	41 (40.6)	0.504
GDM with WHO 1999 (%)	16 (15.5)	9 (8.8)	11 (10.8)	0.307
GDM with WHO 2013 (%)	25 (24.3)	25 (24.5)	23 (22.5)	0.938
Fasting glucose, (mmol/L)	4.38 (0.44)	4.34 (0.56)	4.35 (0.46)	0.764
2-hour glucose, (mmol/L)	5.87 (1.52)	5.49 (1.33)	5.37 (1.42)	0.044
Pregnancy BMI, (kg/m2)	28.12 (5.55)	27.39 (4.41)	27.56 (4.15)	0.521
HOMA-IR	1.45 [1.18, 1.83]	1.52 [1.19, 1.93]	1.42 [1.13, 1.84]	0.858
HOMA-B	175.66 (45.09)	185.31 (53.75)	180.64 (53.54)	0.407
Insulin, (pmol/L)	50.00 [33.00, 71.00]	48.00 [34.00, 67.25]	44.50 [32.00, 73.00]	0.858
C-peptide, (pmol/L)	763.35 (291.45)	795.56 (368.73)	786.49 (364.27)	0.790
Total body fat mass, (kg)	30.38 (11.32)	28.45 (9.20)	29.65 (8.25)	0.365
South Asians	**[0.897 - 0.928]**	**(0.928 - 0.937]**	**(0.937 - 0.954]**	
n	55	55	55	
Age	28.00 (4.29)	28.27 (5.15)	28.36 (4.35)	0.912
Smoking, never (%)	49 (90.7)	53 (96.4)	50 (90.9)	0.526
GDM with WHO 1999 (%)	10 (18.5)	5 (9.3)	9 (16.4)	0.364
GDM with WHO 2013 (%)	20 (36.4)	21 (38.2)	27 (49.1)	0.341
Fasting glucose, (mmol/L)	4.51 (0.43)	4.50 (0.51)	4.46 (0.49)	0.856
2-hour glucose, (mmol/L)	6.09 (1.72)	5.75 (1.72)	6.09 (1.48)	0.472
Pregnancy BMI, (kg/m2)	26.95 (4.01)	27.05 (4.21)	26.59 (4.12)	0.829
HOMA-IR	1.89 [1.57, 2.25]	1.76 [1.44, 2.33]	1.63 [1.36, 2.19]	0.659
HOMA-B	182.44 (33.87)	185.78 (44.65)	192.22 (75.80)	0.638
Insulin, (pmol/L)	74.00 [58.25, 90.00]	75.00 [58.50, 106.00]	68.00 [49.00, 93.75]	0.705
C-peptide, (pmol/L)	904.96 (303.18)	896.55 (326.46)	917.06 (442.76)	0.957
Total body fat mass, (kg)	26.56 (8.34)	26.21 (7.98)	26.05 (8.69)	0.951

The categorical variables are presented as mean ± standard deviation (SD) or n (%), where the chi-squared test were applied. The continuous variables are presented as mean ± SD, where the ANOVA test was applied. The non-normal variables are presented as median, where the Kruskal-Wallis test was applied.

4.3.1 Candidate CpG sites from other studies

Association between previously known candidate CpG sites from EWAS (73-75) and FPG, and 2-h PG was tested using ANOVA (Tables 8 – 9). We found the positive association between FPG and methylation of two CpG sites: cg06500161 and cg11024682 in South Asians and negative association between FPG and cg00574958 methylation in Europeans. The association between cg04999691 and 2-h PG was significant in Europeans. The borderline significant association was observed between 2-h PG and methylation of cg06500161 and cg04999691, and FPG and cg09152259 methylation in South Asians. The negative borderline significant association was observed between 2-h PG and methylation of cg18181703, cg00574958, cg07988171, and FPG and cg19693031 in Europeans.

The applying of Person test indicated some new significant associations: the negative association between cg07988171 methylation with FPG and 2-h PG in Europeans and the positive association between cg11024682 methylation and 2-h PG in South Asians. In addition, three associations that was indicated as borderline significant using ANOVA test, was indicated significant using the Pearson correlation: the negative association between methylation of cg18181703 and cg00574958 with 2-h PG in Europeans, the negative association between the cg19693031 and FPG in Europeans, and the positive association between the cg06500161 methylation and FPG in South Asians.

Table 8: Association between tertiles of DNA methylation at CpG sites previously associated with T2DM (73, 74) and FPG values from STORK Groruddalen cohort. The ANOVA and Pearson correlations were used.

FPG	Low % methylation	Medium % methylation	High % methylation	p	Pearson cor.	Pearson P-value
cg19693031						
European	4.45 (0.55)	4.34 (0.40)	4.29 (0.50)	0.061	-0.150	0.010
South Asians	4.55 (0.45)	4.52 (0.58)	4.40 (0.38)	0.221	-0.086	0.281
cg06500161						
European	4.31 (0.42)	4.34 (0.44)	4.42 (0.59)	0.234	0.076	0.196
South Asians	4.37 (0.42)	4.45 (0.45)	4.65 (0.51)	0.006	0.220	0.005
cg09152259						
European	4.40 (0.56)	4.34 (0.51)	4.32 (0.40)	0.496	-0.038	0.518
South Asians	4.62 (0.50)	4.43 (0.45)	4.43 (0.46)	0.060	-0.068	0.394
cg04999691						
European	4.38 (0.58)	4.39 (0.48)	4.29 (0.39)	0.324	-0.034	0.553
South Asians	4.57 (0.52)	4.44 (0.46)	4.46 (0.44)	0.296	-0.082	0.302
cg11024682						
European	4.32 (0.41)	4.42 (0.46)	4.33 (0.58)	0.308	-0.007	0.904
South Asians	4.36 (0.42)	4.51 (0.48)	4.60 (0.50)	0.026	0.228	0.004
cg02650017						
European	4.31 (0.39)	4.37 (0.54)	4.39 (0.53)	0.478	0.016	0.784
South Asians	4.58 (0.45)	4.49 (0.54)	4.40 (0.42)	0.157	-0.117	0.139
cg18181703						
European	4.41 (0.57)	4.31 (0.44)	4.35 (0.45)	0.352	-0.071	0.221
South Asians	4.49 (0.52)	4.49 (0.45)	4.49 (0.47)	0.998	0.013	0.868
cg04816311						
European	4.39 (0.54)	4.35 (0.45)	4.34 (0.48)	0.753	-0.042	0.472
South Asians	4.47 (0.49)	4.50 (0.53)	4.51 (0.41)	0.894	0.035	0.657
cg00574958						
European	4.46 (0.60)	4.30 (0.38)	4.31 (0.46)	0.044	-0.107	0.065
South Asians	4.57 (0.51)	4.47 (0.43)	4.44 (0.48)	0.345	-0.117	0.139
cg07988171						
European	4.43 (0.61)	4.33 (0.47)	4.31 (0.35)	0.168	-0.135	0.020
South Asians	4.46 (0.48)	4.56 (0.51)	4.44 (0.44)	0.361	0.023	0.770

Table 9: Association between low-medium-high methylation levels of CpG sites associated with T2DM (73, 74) found from other studies and 2-h PG values from STORK Groruddalen cohort. The ANOVA and Pearson correlations were used.

2-h PG	Low % methylation	Medium % methylation	High % methylation	p	Pearson cor.	Pearson P-value
cg19693031						
European	5.63 (1.53)	5.47 (1.33)	5.61 (1.44)	0.706	-0.069	0.245
South Asians	5.93 (1.48)	6.33 (1.88)	5.68 (1.50)	0.119	-0.051	0.526
cg06500161						
European	5.50 (1.57)	5.60 (1.26)	5.62 (1.47)	0.839	0.013	0.828
South Asians	5.55 (1.55)	6.22 (1.61)	6.16 (1.71)	0.071	0.112	0.162
cg09152259						
European	5.60 (1.48)	5.68 (1.52)	5.44 (1.30)	0.489	-0.030	0.607
South Asians	6.21 (1.52)	5.84 (1.68)	5.88 (1.72)	0.449	0.019	0.817
cg04999691						
European	5.72 (1.49)	5.78 (1.42)	5.21 (1.33)	0.011	-0.107	0.071
South Asians	6.31 (1.94)	5.54 (1.20)	6.03 (1.59)	0.050	-0.066	0.416
cg11024682						
European	5.58 (1.33)	5.71 (1.42)	5.42 (1.54)	0.389	-0.055	0.350
South Asians	5.63 (1.42)	6.02 (1.77)	6.26 (1.67)	0.150	0.166	0.0389
cg02650017						
European	5.37 (1.46)	5.59 (1.41)	5.75 (1.42)	0.189	0.081	0.170
South Asians	6.10 (1.66)	5.85 (1.67)	5.97 (1.61)	0.732	-0.030	0.711
cg18181703						
European	5.85 (1.57)	5.46 (1.34)	5.40 (1.35)	0.059	-0.126	0.033
South Asians	6.04 (1.78)	6.12 (1.54)	5.73 (1.58)	0.463	-0.039	0.630
cg04816311						
European	5.67 (1.38)	5.65 (1.41)	5.40 (1.51)	0.360	-0.065	0.271
South Asians	6.14 (1.70)	5.98 (1.70)	5.80 (1.53)	0.573	-0.103	0.200
cg00574958						
European	5.83 (1.49)	5.53 (1.37)	5.35 (1.41)	0.058	-0.119	0.043
South Asians	6.22 (1.80)	6.01 (1.40)	5.70 (1.70)	0.271	-0.066	0.417
cg07988171						
European	5.82 (1.52)	5.52 (1.44)	5.38 (1.31)	0.099	-0.137	0.02
South Asians	5.91 (1.33)	5.71 (1.53)	6.32 (1.99)	0.156	0.143	0.076

We did not find significant results associated with cg06500161 methylation in Europeans, but the results from South Asians seems to be statistically significant both for insulin (with p-value 0.003) and HOMA-IR (p-value 0.002) (Table 10).

Table 10: Association between tertiles of DNA methylation at cg06500161 and insulin and HOMA-R-values in Europeans and South Asians from STORK Groruddalen cohort.

Europeans	**[0.519 - 0.594]**	**[0.594 - 0.614]**	**[0.614 - 0.693]**	**p (1)**	**rho**	**p (2)**
Insulin, (pmol/L)	46.00 [33.00, 65.00]	47.50 [34.75, 69.25]	54.00 [33.00, 75.75]	0.199	0.111	0.053
HOMA-IR	1.57 (0.71)	1.64 (0.76)	1.73 (0.78)	0.338	0.115	0.045
South Asians	**[0.556 - 0.602]**	**[0.602 - 0.622]**	**[0.622 - 0.68]**			
Insulin, (pmol/L)	63.00 [38.50, 86.50]	71.00 [49.00, 86.00]	82.00 [62.50, 121.50]	0.003	0.289	<0.001
HOMA-IR	1.66 (0.64)	1.95 (0.81)	2.18 (0.84)	0.002	0.289	<0.001

The Kruskal-Wallis (p (1)) and Spearman (p (2)) tests were applied.

5 Discussion

5.1 Methodological considerations

5.1.1 Study design

The STORK Groruddalen project is a prospective population-based cohort study where pregnant women were followed from early gestation up to tree mouths postpartum. The advantages of cohort studies are the possibilities for measuring exposure before the disease manifestation and to test for multiple outcomes (103, p.36). However, like others observational research, cohort studies can only give clues about associations, and not evidence of causality.

The impossibility of causality determination in observational studies is due to confounding and bias. Additional experimental study may minimize confounding problems, but this design cannot always be applied. However, the causality of observed association may be assumed by satisfying the Hill's criteria for causation (104). This master thesis had a cross-sectional study design. The challenges with the cross-sectional study design is satisfying of the Hill's criteria with temporal relationship of the observed association in the study. The design give detailed description of the population at a certain time without the possibility of looking at exposures over time or to test if the exposure or outcome happens first. Nevertheless, those studies may give new hypotheses that may be further tested.

5.1.2 Internal validity

Internal validity represent the accuracy of the study. High validity refers to strong research methods minimizing confounding and other bias (105). Selection bias, confounding and information bias will affect the internal validity.

Selection bias

Selection bias occurs when the examined cohort does not represent the population of interest (106). Systematic refusal to participating in the study of certain pregnant women with some unifying variables may result in this type of bias. Lower participation rate may imply

higher chance for selection bias (103, p.94). However, the participation rate in STORK Groruddalen was high with no observed significant differences between included and excluded women (76).

Confounding

A confounding factor is a variable related both to the exposure and to outcome, but is not an intermediate between those (107). We attempted to identify potential confounding factors by looking at associations between candidate CpG sites and various phenotypes. However, we can never be sure we have excluded all confounding factors in observational studies due to potential unknown or unmeasured confounders.

Information bias

Information bias is systematic measurement errors during the data collection (108).

EthnicityThe ethnicity of participating women was determined based on genetic genome-wide patterns from a genome-wide association study in the sample (Moen et al. unpublished). In this way, although non-European women who are adopted may self-report their ethnicity as ethnic Norwegian, the ethnic affiliation used in the study will be based on genetics instead. Although genetics are thought to influence DNA methylation patterns, originating from a geographic area often imply belonging to a specific culture with specific lifestyle that may influence epigenetic and thereby affect the results. However, most of the women had the same genetically ethnicity as self-reported.

Plasma glucose samplesThe HemoCue method was used for glucose measuring and GDM diagnosis. The plasma glucose was measured on site without shipping to laboratory. This eliminated waiting time and glycolysis was avoided, giving more accurate glucose values (109). Despite this advantage, the method may give falsely low glucose levels in presence of high triglyceride levels (110).

White blood cells

The methylation analysis was performed on white blood cells. The main reason for choosing these cells was easy accessibility. However, there are many possible research topics for understanding the pathogenesis of hyperglycemia in pregnancy (111), and the best tissues to

study may vary. To study how DNA methylation alteration affect the insulin secretion, the perfect tissue would probably be pancreatic β cells. Likewise, muscle cells, liver and adipose tissue may be helpful tissues for understanding how DNA methylation may affect development of IR. Unfortunately, these tissues are not readily available due to necessity of surgical intervention and the following risk for the patient. In addition, the increasing evidence confirm that epigenetic alterations may be observed also in leukocytes and is not limited to specific affected tissues (112).

The whole blood samples include different cell types, but only white blood cells have nucleus with DNA. The cell composition of white blood cells include lymphocytes, neutrophils, monocytes, basophils and eosinophils. Each of them have different methylation patterns (113), which can lead to bias in the results if the cell composition is not accounted for. The EWAS analyses used to pick candidate CpG sites were adjusted for cell composition to avoid this bias, while the β values studied in this master thesis were not, which may explain some differences in the results. Despite this, it will not affect the EPIC validation results. The methylation values from BSP method were not adjusted for cell composition in the same way as β values from the EPIC and both methods will give similar values.

DNA methylation analysis

The Illumina Methylation EPIC BeadChip is a new chip produced by Illumina in addition to the previous version, the HumanMethylation450 BeadChip (450K). The 450K have possibility to measure DNA methylation of 450,000 CpG sites, while EPIC BeadChip covers over 850,000 CpG sites and give a more complete picture of human methylome (114).

The EPIC BeadChip was technically validated by BSP. Both methods starts with DNA bisulfite conversion. Incomplete conversion may result in false positive bias: More cytosines will be indicated as methylated. Further, both methods provide DNA amplification. EPIC uses whole genome amplification (WGA), while BSP apply by using PCR. The WGA procedure take more time (overnight) comparing to PCR (2 hours) and results in fewer DNA copies number (several thousands) that PCR (several trillions) (115). Therefore, the DNA input is higher in WGA (250 ng) than in PCR (20-40 ng) (116). Thereby, the BSP may presuppose PCR bias: Methylated and unmethylated DNA templates seems to be PCR amplified with different efficiency, favoring unmethylated DNA amplification (117). In cases with low frequency of methylated templates, the methylation status can be underestimated. In addition,

low quantity of PCR DNA input with further high amplification have more chance for bias. Higher amount of DNA input as in WGA with fewer amplification copies will represent more correct distribution of differently methylated DNA molecules.

The EPIC is susceptible to technical confounding as batch effect (118, 119), where the results can vary with each array run. To minimize this confounding, the β values from the EPIC were normalized for batch effect. The normalization can reduce the effect, but not avoid it completely (118). It can be many reasons for the effect: technical tools, different sample plates, relatively long time of procedure and different processing days, differences in laboratory personal work and general laboratory conditions. The BSP can also be influenced by the batch effect in some degree, but less due to a shorter protocol.

The BSP performed in this thesis was performed in parallels and up to seven times for some samples. The variation between methylation values obtained from each pyrosequencing run varied in average between 0.91 - 1.44 % (Appendix I, Tables 11 – 13), which indicate high reproducibility of the method.

The β values in the EPIC are retrieved from either Infinium I or Infinium II assays. The Infinium I and II assays seem to be diverged and have different distributions (119, 120). Furthermore, the β values from Infinium II assay are less accurate compared to Infinium I. The inexact results from both assays in the same array may lead to a signal shift and misbalance during results analysis. Currently there are several methods developed to normalize the β values from Infinium II assay (119, 121-123).

Despite some bias in the EPIC, the 450K chip, which the EPIC is based on, is highly accurate and may be used for genome methylation analysis with high reproducibility (119).

5.1.3 External validity

The external validity indicate the ability to transfer the study results to other populations. With other words, it is generalizability of the study. The strength of the study is the high attendance. The ethnic diversity and education level presented in STORK Groruddalen study is quite representative for ethnic and education level distribution in Groruddalen area, the area with higher number of ethnic minorities than otherwise in Oslo and in Norway in general, and lower education level. In this way, the results of the study can be applied to other healthy

women living in Groruddalen or other districts with similar high amount of non-Western ethnic background (76). The mean age of participants in the study is comparable to the average childbearing age of women in Norway which was 31.7 years in 2008-2010 (124).

5.2 Main findings

5.2.1 Technical validation

We observed a strong correlation between BSP and the EPIC. The average difference in methylation percentage found by BSP and by EPIC was different for each CpG site. The differences in cg17148978 and cg08098128 was less than 1 % (0.78 % and 0.84 % respectively) (Appendix I, Tables 11 and 13), which indicate high precision. However, the difference in cg14120215 is average 8.7 % (Appendix I, table 12).

Our r-value for the overall correlation was lower than r-values observed in other studies by technical validation of 450K (r-values $\geq$ 0.89 (125-127)). The results from 450K and EPIC are thought to be comparable because both use Infinium I and Infinium II assays, however more validation analysis are needed, since the EPIC include much more CpG sites tested on Infinium II assays.

The β values from the EPIC was not normalized in regards to Infinium II assay. At the same time, the all three BSP analyzed CpG sites were from Infinium II assay and lack of normalization may affect the correlation analysis.

The possible reason for some variation between BSP and the EPIC may be an absence of normalization for Infinium II assay or other technical bias. Further normalization for Infinium II assay may increase the correlation. In addition, validating another CpG site with higher average methylation level could be important to see how results varies with increased methylation. For this reason, BSP of the CpG site cg19327414 with methylation level 90 - 95 % is ongoing.

5.2.2 Candidate CpG sites

Exploring the tree CpG sites from EPIPREG study in tertiles indicated significant association with BMI (cg08098128), age (cg14120215) and 2-h PG (cg19327414).

Possible confounding

The CpG site cg08098128 was found to be significant associated with FPG in Europeans in the EPIPREG EWAS, while the significant association was not observed in tertiles. However, we observed a positive trend with FPG within the tertiles. We found a significant positive association between BMI and cg08098128 in the Europeans. Since the EWAS was adjusted for BMI, but values from tertiles not, the BMI is probably a confounding factor.

The CpG site cg14120215 was found to be also associated with FPG in Europeans in the EPIPREG EWAS. Nevertheless, we did not find a similar association for FPG values in tertiles. The lack of significant association between DNA methylation and FPG in tertiles may be due to that the EWAS results were adjusted to BMI, age, average cell composition and smoking status, while the statistical analysis in tertiles was performed with raw data without adjusting. The different results between the EPIPREG EWAS and tertiles results may indicate that these factors may confound the results. A significant negative association was found with age in Europeans (Table 6). Since age is associated with GDM (16), age could be a confounding factor, explaining why we did not find an association between the DNA methylation. In addition, there was found association with 2-h PG, where the association was u-shaped.

In spite of lack of adjusting results in the tertiles analysis, we found a significant negative association of cg19327414 methylation with 2-h PG. We did not observe any other associations, either in Europeans or in South Asians, suggesting absence of potential confounders.

5.2.1 Possible biological mechanisms

cg08098128

The cg08098128 is located in the *STK16* gene which code for the serine/threonine kinase 16. The protein phosphorylates serine and threonine (128). Increased FPG is associated with increased methylation of the CpG, which lead to lower translation of the protein. The protein has an important role in regulation of Golgi apparatus (129) (130) and probably important in intracellular vesicular traffic. In addition, the kinase may be involved in TGF-β signaling (131). The *STK16* gene is not known to be associated with T2DM, GDM or

increased plasma glucose in previous genetic or epigenetic studies, but further studying of the STK16 protein function may be useful for understanding of the association with increased FPG.

cg14120215

The CpG site cg14120215 is located in the *SDS* gene. The gene codes for L-serine dehydratase/L-threonine deaminase (SDH), which is involved in serine, threonine and glycine metabolism (132). The enzyme is involved in the gluconeogenesis pathway. The highest expression of the protein is in the liver (133). An experiment performed in rat's liver indicated that increased glucose and insulin levels inhibit the SDH induction (activity and amount of enzyme) (134). In our study, a positive correlation between the CpG methylation and FPG was found. One of the disadvantages of the cross-sectional study is difficulty of determination what cause is and what is an outcome - if the CpG methylation affect FPG levels or if increased FPG lead to increased DNA methylation. The study on rat's liver may indicate that increased FPG values lead to increased CpG methylation and further inhibition of SDH induction.

In addition, an association between SDH and the diabetic retinopathy pathogenesis has been indicated. The diabetic retinopathy is a vascular disease, where retina blood vessels damages due to T2DM complication and may result in partial or complete vision loss (135). Hence, we can speculate that DNA methylation due to increased FPG is involved in development of T2DM complications.

cg19327414

The CpG site cg19327414 is located near the *PNLDC1* gene, which code for poly(A)-specific ribonuclease PNLDC1 (136). The ribonuclease cleaves the poly(A) tails of mRNA, performing the first step in mRNA decay. One study found that the protein is expressed in early development and during cell differentiation to provide clearance of maternal mRNA (137). In addition, the protein is found to be associated with affective psychoses (138) and chronic kidney disease (139). Some studies found an association between hyperglycemia and T2DM with chronic kidney disease (140-142), suggesting that methylation of this CpG site play a role in development of T2DM complication as chronic kidney disease due to increased 2-h PG.

CpG sites from other studies

None of the CpG sites discovered in the EPIPREG study (Moen et al, unpublished) has been found by published EWAS. However, EWAS results are often not comparable due to differences across samples. Also, other phenotypes could confound the relationships if they are associated with both DNA methylation and diabetes, such as chronic kidney disease (143), liver cancer (144), vascular diseases (145, 146) and Alzheimer's disease (147). Those diseases share some risk factors and pathogenetic mechanisms with T2DM that actually could influence the result from other studies. In addition, the previously published EWAS from other groups have all adjusted for age and sex (74, 75), but only one study adjusted for age and sex, cell counts and BMI (73). Furthermore, since some of the previously published EWAS did not adjust for cell composition and BMI, this may be a reason why we did not observe the same associations in our cohort. Also, the participants in EPIPREG study were young women, while other studies included both men and women with different age and most of diabetic patients are older. Those factors can affect methylation and affect the results.

Ethnicity

The association between plasma glucose and the tree CpG sites from the EPIPREG study was significant only in European group. The DNA methylation tertiles indicated significant association with BMI, age and 2-h PG only in Europeans, while no association was observed in South Asian group, suggesting that it might be methylation differences across the ethnic groups (74, 148, 149).

Similarly, the association between plasma glucose values and methylation of previously known CpG sites from published studies were significant in four of ten CpG sites using ANOVA: cg06500161 (SA, FPG), cg11024682 (SA, FPG), cg00574958 (EU, FPG) and cg04999691 (EU, 2-h PG). In addition, seven CpG sites were borderline significantly associated with plasma glucose. While, with Pearson correlation, the significant association was found in six CpG sites: cg19693031 (EU, FPG), cg06500161 (SA, FPG), cg11024682 (SA, FPG and SA, 2-h PG), cg07988171 (EU, FPG and EU, 2-h PG), cg18181703 (EU, 2-h PG), cg00574958 (EU, 2-h PG) and one CpG site was borderline significantly associated with 2-h PG. The associations varied between Europeans and South Asians - again suggesting ethnic differences.

These differences could be a result of different life styles between the Europeans and South Asians. Life style include different food preference (150), lower physical activity of South Asians (151), lower alcohol-intake before pregnancy in South Asians, higher smoking status in Europeans (Table 4) (149), social-economic status (Table 4) and others factors, as well as genetic differences which could lead to differences in methylation patterns. Additionally, some CpG methylation can be maintained throughout generations (152). The methylations pattern in this way can be inherited and tied to a specific ethnic group.

Also, age is known to affect methylation patterns (153) and the South Asians were 2 years younger that the European women.

Statistical power

The use of tertiles result in lower statistical power in our statistical analyses. Applying the Pearson correlation for correlation identified a larger number of significant associations.

The sample sizes of Europeans and especially South Asians in this master thesis are quite low. This may be a reason for why the CpG sites from EPIPREG study was observed only in European cohort and there was not observed any CpG sites associated with plasma glucose in South Asians. Low statistical power may also explain why the CpG sites found in other studies were not significant observed in this master study. These studies had 713 participants (73), 837 participants (75) and 2664 participants (74), while the EPIPREG study have less participants (n=480). A higher number of participants would increase our statistical power, but would not have been feasible due to the high costs of the EPIC.

6 Conclusions

- We observed a strong correlation between the Illumina MethylationEPIC BeadChip and BSP suggesting high technical validity of the chip. However, further normalization of β values from Infinium II assay may probably give values that are more correct and thereby indicate stronger correlation.

- The positive association between cg08098128 methylation (associated with FPG) and BMI in Europeans was found in tertiles. In the same way, the negative association was found between cg14120215 methylation (associated with FPG) and age in Europeans. In this way, the BMI and age are confounding factors, but EWAS was adjusting for those will minimize the effect of this. For the same reason, as adjusting for different factors of EWAS results, we did not observed significant association between those two CpG sites and FPG. However, there was observed significant association between cg19327414 methylation and 2-h PG in Europeans and no obvious confounding was observed for this CpG site.

- The three CpG sites from the EPIPREG EWAS of plasma glucose studied in this master thesis are located in different positions in the genome and affect the transcription of different genes. Weakened transcription may affect the production of proteins translated from those genes. Future studies should explore how these proteins are associated with increased plasma glucose values or possibly with GDM or T2DM.

- Four associations between methylation of CpG sites found from other EWAS studies associated with T2DM was found to be associated with plasma glucose in the EPIPREG sample, both in South Asians and Europeans, using ANOVA. The applying of Person correlation indicated larger number of significant associations between CpG methylation and plasma glucose levels. Not all CpG sites confirmed the association with increased glucose plasma levels.

References

1. Hadden DR, McLaughlin C. Normal and abnormal maternal metabolism during pregnancy. Seminars in fetal & neonatal medicine. 2009;14(2):66-71.
2. Battaglia FC. Principal substrates of fetal metabolism: fuel and growth requirements of the ovine fetus. Ciba Foundation symposium. 1978(63):57-74.
3. Lain KY, Catalano PM. Metabolic changes in pregnancy. Clinical obstetrics and gynecology. 2007;50(4):938-48.
4. Newbern D, Freemark M. Placental hormones and the control of maternal metabolism and fetal growth. Current opinion in endocrinology, diabetes, and obesity. 2011;18(6):409-16.
5. Sivan E, Homko CJ, Chen X, Reece EA, Boden G. Effect of insulin on fat metabolism during and after normal pregnancy. Diabetes. 1999;48(4):834-8.
6. Williams DA, Lemke TL. Foye's principles of medicinal chemistry. 7th ed. ed. Philadelphia: Lippincot Williams & Wilkins; 2013.
7. Sand O, Haug E, Sjaastad ØV, Toverud KC. Menneskets fysiologi. 1. utg, 1. oppl. ed. Oslo: Gyldendal akademisk; 2001.
8. Saltiel AR, Kahn CR. Insulin signalling and the regulation of glucose and lipid metabolism. Nature. 2001;414(6865):799-806.
9. Sonagra AD, Biradar SM, K D, Murthy DSJ. Normal pregnancy- a state of insulin resistance. J Clin Diagn Res. 2014;8(11):CC01-3.
10. Barbour LA, McCurdy CE, Hernandez TL, Kirwan JP, Catalano PM, Friedman JE. Cellular mechanisms for insulin resistance in normal pregnancy and gestational diabetes. Diabetes care. 2007;30 Suppl 2:S112-9.
11. Bajaj M, Defronzo RA. Metabolic and molecular basis of insulin resistance. Journal of nuclear cardiology : official publication of the American Society of Nuclear Cardiology. 2003;10(3):311-23.
12. Buchanan TA, Xiang AH. Gestational diabetes mellitus. The Journal of clinical investigation. 2005;115(3):485-91.
13. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, et al. Hyperglycemia and adverse pregnancy outcomes. The New England journal of medicine. 2008;358(19):1991-2002.
14. Moen GH, Sommer C, Prasad RB, Sletner L, Groop L, Qvigstad E, et al. MECHANISMS IN ENDOCRINOLOGY: Epigenetic modifications and gestational diabetes: a systematic review of published literature. European journal of endocrinology. 2017;176(5):R247-r67.
15. Williams MA, Qiu C, Dempsey JC, Luthy DA. Familial aggregation of type 2 diabetes and chronic hypertension in women with gestational diabetes mellitus. The Journal of reproductive medicine. 2003;48(12):955-62.
16. Teh WT, Teede HJ, Paul E, Harrison CL, Wallace EM, Allan C. Risk factors for gestational diabetes mellitus: implications for the application of screening guidelines. The Australian & New Zealand journal of obstetrics & gynaecology. 2011;51(1):26-30.
17. Jenum AK, Sommer C, Sletner L, Morkrid K, Baerug A, Mosdol A. Adiposity and hyperglycaemia in pregnancy and related health outcomes in European ethnic minorities of Asian and African origin: a review. Food & nutrition research. 2013;57.
18. Sommer C, Morkrid K, Jenum AK, Sletner L, Mosdol A, Birkeland KI. Weight gain, total fat gain and regional fat gain during pregnancy and the association with gestational diabetes: a population-based cohort study. International journal of obesity (2005). 2014;38(1):76-81.
19. Xiong X, Saunders LD, Wang FL, Demianczuk NN. Gestational diabetes mellitus: prevalence, risk factors, maternal and infant outcomes. [New York] :2001. p. 221-8.
20. Helsedirektoratet. Diagnostikk og tiltak for å finne uoppdaget diabetes og svangerskapsdiabetes. : Helsedirektoratet; [cited 2018 11.06]. Available from: https://helsedirektoratet.no/retningslinjer/svangerskapsdiabetes/seksjon?Tittel=diagnostikk-og-tiltak-for-20014570.

21. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes care. 2002;25(10):1862-8.
22. Dornhorst A, Frost G. The potential for dietary intervention postpartum in women with gestational diabetes. Am Diabetes Assoc; 1997.
23. Giacco F, Brownlee M. Oxidative stress and diabetic complications. Circulation research. 2010;107(9):1058-70.
24. Fowler MJ. Microvascular and macrovascular complications of diabetes. Clinical diabetes. 2008;26(2):77-82.
25. Birkeland KI. Diabetes mellitus. In: Birkeland KI, Gullestad L, Aabakken L, editors. Indremedisin I & II. Drammen: Vett & Viten 2017. p. 197-218.
26. Coustan DR, Lowe LP, Metzger BE, Dyer AR. The Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study: paving the way for new diagnostic criteria for gestational diabetes mellitus. American journal of obstetrics and gynecology. 2010;202(6):654.e1-6.
27. Damm P, Houshmand-Oeregaard A, Kelstrup L, Lauenborg J, Mathiesen ER, Clausen TD. Gestational diabetes mellitus and long-term consequences for mother and offspring: a view from Denmark. Diabetologia. 2016;59(7):1396-9.
28. Challier JC, Hauguel S, Desmaizieres V. Effect of insulin on glucose uptake and metabolism in the human placenta. The Journal of clinical endocrinology and metabolism. 1986;62(5):803-7.
29. Kc K, Shakya S, Zhang H. Gestational diabetes mellitus and macrosomia: a literature review. Annals of nutrition & metabolism. 2015;66 Suppl 2:14-20.
30. Catalano PM, Hauguel-De Mouzon S. Is it time to revisit the Pedersen hypothesis in the face of the obesity epidemic? American journal of obstetrics and gynecology. 2011;204(6):479-87.
31. Fowden AL. The role of insulin in fetal growth. Early human development. 1992;29(1-3):177-81.
32. Jovanovic L. The role of continuous glucose monitoring in gestational diabetes mellitus. Diabetes technology & therapeutics. 2000;2 Suppl 1:S67-71.
33. Arya VB, Senniappan S, Guemes M, Hussain K. Neonatal hypoglycemia. Indian journal of pediatrics. 2014;81(1):58-65.
34. Clausen TD, Mathiesen ER, Hansen T, Pedersen O, Jensen DM, Lauenborg J, et al. Overweight and the metabolic syndrome in adult offspring of women with diet-treated gestational diabetes mellitus or type 1 diabetes. The Journal of clinical endocrinology and metabolism. 2009;94(7):2464-70.
35. Dabelea D. The predisposition to obesity and diabetes in offspring of diabetic mothers. Diabetes care. 2007;30 Suppl 2:S169-74.
36. Lauenborg J, Grarup N, Damm P, Borch-Johnsen K, Jørgensen T, Pedersen O, et al. Common type 2 diabetes risk gene variants associate with gestational diabetes. 2009;94(1):145-50.
37. Cho Y, Kim T, Lim S, Choi S, Shin H, Lee H, et al. Type 2 diabetes-associated genetic variants discovered in the recent genome-wide association studies are related to gestational diabetes mellitus in the Korean population. 2009;52(2):253-61.
38. Pappa KI, Gazouli M, Economou K, Daskalakis G, Anastasiou E, Anagnou NP, et al. Gestational diabetes mellitus shares polymorphisms of genes associated with insulin resistance and type 2 diabetes in the Greek population. 2011;27(4):267-72.
39. Robitaille J, Grant AM. The genetics of gestational diabetes mellitus: evidence for relationship with type 2 diabetes mellitus. Genetics in medicine : official journal of the American College of Medical Genetics. 2008;10(4):240-50.
40. Ling C, Groop L. Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes. 2009;58(12):2718-25.
41. Krebs JE, Goldstein ES, Kilpatrick ST. Lewin's genes XII. 12. ed: Jones & Bartlett Learning; 2018.
42. Pal S, Tyler JK. Epigenetics and aging. Science advances. 2016;2(7):e1600584.

43. Hackett JA, Surani MA. DNA methylation dynamics during the mammalian life cycle. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2013;368(1609):20110328.
44. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. Nature reviews Genetics. 2012;13(2):97-109.
45. Watson JD. Molecular biology of the gene. 7th ed. ed. Boston: Pearson; 2014.
46. Alberts B, Wilson JH, Hunt T. Molecular biology of the cell. 6th ed. ed. New York: Garland Science; 2015.
47. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693-705.
48. Becker PB, Workman JL. Nucleosome remodeling and epigenetics. Cold Spring Harbor perspectives in biology. 2013;5(9).
49. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell. 1999;98(3):285-94.
50. Wheeler R. Nucleosome structure-2: Wikimedia Commons 2011 [cited 2018 13.08]. Schematic representation of the assembly of the core histones into the nucleosome]. Available from: https://commons.wikimedia.org/wiki/File:Nucleosome_structure-2.png.
51. Talbert PB, Henikoff S. Histone variants--ancient wrap artists of the epigenome. Nature reviews Molecular cell biology. 2010;11(4):264-75.
52. Maze I, Noh KM, Soshnev AA, Allis CD. Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nature reviews Genetics. 2014;15(4):259-71.
53. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell research. 2011;21(3):381-95.
54. Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, et al. S-Adenosylmethionine and methylation. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 1996;10(4):471-80.
55. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews Genetics. 2012;13(7):484-92.
56. Ahmed H. Promoter Methylation in Prostate Cancer and its Application for the Early Detection of Prostate Cancer Using Serum and Urine Samples. Biomarkers in cancer. 2010;2010(2):17-33.
57. Razin A, Riggs AD. DNA methylation and gene function. Science (New York, NY). 1980;210(4470):604-10.
58. Bird A. DNA methylation patterns and epigenetic memory. Genes & development. 2002;16(1):6-21.
59. Craig JM, Wong NC. Epigenetics : a reference manual. Norfolk: Caister Academic Press; 2011.
60. Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. Science (New York, NY). 2001;293(5532):1068-70.
61. Wade PA. Methyl CpG-binding proteins and transcriptional repression. BioEssays : news and reviews in molecular, cellular and developmental biology. 2001;23(12):1131-7.
62. Eden S, Hashimshony T, Keshet I, Cedar H, Thorne AW. DNA methylation models histone acetylation. Nature. 1998;394(6696):842.
63. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PloS one. 2013;8(5):e63812.
64. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, et al. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. Human molecular genetics. 2012;21(13):3073-82.
65. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. Frontiers in genetics. 2013;4:132.
66. Zhang N. Epigenetic modulation of DNA methylation by nutrition and its mechanisms in animals. Animal nutrition (Zhongguo xu mu shou yi xue hui). 2015;1(3):144-51.

67. Jung M, Pfeifer GPJBB. Aging and DNA methylation. 2015;13(1):7.
68. Zhang FF, Cardarelli R, Carroll J, Zhang S, Fulda KG, Gonzalez K, et al. Physical activity and global genomic DNA methylation in a cancer-free population. Epigenetics. 2011;6(3):293-9.
69. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nature reviews Genetics. 2011;12(8):529-41.
70. Wu P, Farrell WE, Haworth KE, Emes RD, Kitchen MO, Glossop JR, et al. Maternal genome-wide DNA methylation profiling in gestational diabetes shows distinctive disease-associated changes relative to matched healthy pregnancies. Epigenetics. 2018;13(2):122-8.
71. Kang J, Lee CN, Li HY, Hsu KH, Lin SY. Genome-wide DNA methylation variation in maternal and cord blood of gestational diabetes population. Diabetes research and clinical practice. 2017;132:127-36.
72. UCSC_Genome_Browser. UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly 2009 [Available from: https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr12%3A113841668-113841668&hgsid=726697349_52g2dUDvdNJxLs0KD86As3PDzAin.
73. Meeks KAC, Henneman P, Venema A, Addo J, Bahendeka S, Burr T, et al. Epigenome-wide association study in whole blood on type 2 diabetes among sub-Saharan African individuals: findings from the RODAM study. International journal of epidemiology. 2018.
74. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. The lancet Diabetes & endocrinology. 2015;3(7):526-34.
75. Hidalgo B, Irvin MR, Sha J, Zhi D, Aslibekyan S, Absher D, 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):801-7.
76. Jenum AK, Sletner L, Voldner N, Vangen S, Morkrid K, Andersen LF, et al. The STORK Groruddalen research programme: A population-based cohort study of gestational diabetes, physical activity, and obesity in pregnancy in a multiethnic population. Rationale, methods, study population, and participation rates. Scandinavian journal of public health. 2010;38(5 Suppl):60-70.
77. Wiggen KS, Dzamarija MT, Thorsdalen B, Østby L. Innvandreres demografi og levekår i Groruddalen, Søndre Nordstrand, Gamle Oslo og Grünerløkka. Oslo–Kongsvinger; 2015.
78. www.publicdomainpictures.net. Беременные, силуэт, леди, мать, [Available from: https://www.publicdomainpictures.net/ru/view-image.php?image=265295&picture=pregnant-silhouette-lady-mother.
79. Silh S. SVG > bird girl pregnancy delivery.
80. publicdomainpictures.net. Mommy And Stroller.
81. Sciences PL. The Complete DELFIA immunodiagnostics system ®.
82. Mørkrid K, Jenum AK, Sletner L, Vårdal MH, Waage CW, Nakstad B, et al. Failure to increase insulin secretory capacity during pregnancy-induced insulin resistance is associated with ethnicity and gestational diabetes. European journal of endocrinology. 2012:EJE-12-0452.
83. Sommer C. Maternal adiposity: Associations with gestational diabetes and neonatal fat in a multi-ethnic population. 2015.
84. Phillips PJ. Oral glucose tolerance testing. Australian family physician. 2012;41(6):391-3.
85. Rutle O, Bruusgaard D, Furuseth K, Vaaler S. Oral glucose tolerance test in general practice--when is it worthwhile? Scandinavian journal of primary health care. 1994;12(4):255-60.
86. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabetic medicine : a journal of the British Diabetic Association. 1998;15(7):539-53.
87. Jenum AK, Mørkrid K, Sletner L, Vange S, Torper JL, Nakstad B, et al. Impact of ethnicity on gestational diabetes identified with the WHO and the modified International Association of Diabetes

and Pregnancy Study Groups criteria: a population-based cohort study. European journal of endocrinology. 2012;166(2):317-24.
88. Oslo universitetssykehus. PF 14 Ekstraksjon av DNA, saltfellingsmetode, Hlab. 2017.
89. Illumina. Infinium® MethylationEPIC BeadChip. 2017.
90. Illumina. Understanding epigenetic changes. Quantitative measurement of methylation at individual CpG sites for high-resolution analysis. [Available from: https://emea.illumina.com/science/technology/beadarray-technology/infinium-methylation-assay.html?langsel=/no/.
91. Illumina. Infinium HD FFPE Methylation Assay, Manual Protocol. 2012.
92. Illumina. Infinium II Assay Multi-Sample protocols.
93. Lu P. Illumina methylation workflow: Wikimedia Commons; 2009 [cited 2017 27.08]. Work flow of Illumina Methylation Assay using the Infinium II platform.]. Available from: https://commons.wikimedia.org/wiki/File:Illuminamethylationworkflow.png.
94. Bassil CF, Huang Z, Murphy SK. Bisulfite pyrosequencing. Methods in molecular biology (Clifton, NJ). 2013;1049:95-107.
95. Qiagen. EpiTect® Bisulfite Handbook. 2014.
96. Qiagen. EpiTect® Whole Bisulfitome Handbook. 2008.
97. Inc. TFS. Quant-iT™ dsDNA Broad-Range Assay Kit 2015 [Available from: https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQuant_iT_dsDNA_BR_Assay_UG.pdf&title=VXNlciBHdWlkZTogUXVhbnQtaVQgZHNETkEgQnJvYWQtUmFuZ2UgQXNzYXkgS2l0.
98. Qiagen. PyroMark® Q48 Autoprep User Manual. 2016.
99. Qiagen. PyroMark Buffers and Solutions 2019 [Available from: https://www.qiagen.com/no/service-and-support/learning-hub/technologies-and-research-topics/pyrosequencing-resource-center/technology-overview/pyromark-buffers-and-solutions/?Print=1.
100. Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. Science (New York, NY). 1998;281(5375):363, 5.
101. Ravilious GE, Herrmann J, Goo Lee S, Westfall CS, Jez JM. Kinetic mechanism of the dimeric ATP sulfurylase from plants. Bioscience reports. 2013;33(4).
102. QIAGEN®. PyroMark® Q48 Autoprep Software. 2015
103. Per Magnus LSB. Epidemiologi: Gyldendal 2013.
104. Hill AB. THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? Proceedings of the Royal Society of Medicine. 1965;58:295-300.
105. McDermott RJChoeps. Internal and external validity. 2011:27-40.
106. Delgado-Rodriguez M, Llorca JJJoE, Health C. Bias. 2004;58(8):635-41.
107. Axelson OJSjweh. Aspects on confounding in occupational health epidemiology. 1978;4(1):98-102.
108. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. Journal of multidisciplinary healthcare. 2016;9:211-7.
109. Wiener K. An Assessment of the Effect of Haematocrit on the Hemocue Blood Glucose Analyser. 1993;30(1):90-3.
110. Ashworth L, Gibb I, Alberti KJCc. HemoCue: evaluation of a portable photometric system for determining glucose in whole blood. 1992;38(8):1479-82.
111. Zhou Z, Sun B, Li X, Zhu C. DNA methylation landscapes in the pathogenesis of type 2 diabetes mellitus. Nutrition & metabolism. 2018;15:47.
112. Pidsley R, Mill J. Epigenetic studies of psychosis: current findings, methodological approaches, and implications for postmortem research. Biological psychiatry. 2011;69(2):146-56.
113. Accomando WP, Wiencke JK, Houseman EA, Nelson HH, Kelsey KT. Quantitative reconstruction of leukocyte subsets using DNA methylation. Genome biology. 2014;15(3):R50.

114. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. 2016;17(1):208.
115. Illiumina. Infinium® HD Assay Methylation Protocol Guide 2015 [Available from: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/infinium_assays/infinium_hd_methylation/infinium-hd-methylation-guide-15019519-01.pdf.
116. Illumina. Infinium MethylationEPIC Kit 2019 [Available from: https://www.illumina.com/products/by-type/microarray-kits/infinium-methylation-epic.html.
117. Wojdacz TK, Borgbo T, Hansen LL. Primer design versus PCR bias in methylation independent PCR amplifications. Epigenetics. 2009;4(4):231-4.
118. Sun Z, Chai HS, Wu Y, White WM, Donkena KV, Klein CJ, et al. Batch effect correction for genome-wide methylation data with Illumina Infinium platform. 2011;4(1):84.
119. Touleimat N, Tost J. Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. 2012;4(3):325-41.
120. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. 2011;3(6):771-84.
121. Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. Bioinformatics. 2016;32(17):2659-63.
122. Maksimovic J, Gordon L, Oshlack AJGB. SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. 2012;13(6):R44.
123. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2012;29(2):189-96.
124. SSB. Statistikkbanken "Fødte" [Available from: https://www.ssb.no/statbank/table/05530/tableViewLayout1/.
125. Glossop JR, Nixon NB, Emes RD, Haworth KE, Packham JC, Dawes PT, et al. Epigenome-wide profiling identifies significant differences in DNA methylation between matched-pairs of T- and B-lymphocytes from healthy individuals. Epigenetics. 2013;8(11):1188-97.
126. Glossop JR, Glossop JR, Emes RD, Emes RD, Nixon NB, Nixon NB, et al. Genome-wide DNA methylation profiling in rheumatoid arthritis identifies disease-associated methylation changes that are distinct to individual T- and B-lymphocyte populations. Epigenetics. 2014;9(9):1228-37.
127. de la Rica L, Urquiza JM, Gómez-Cabrero D, Islam AB, López-Bigas N, Tegnér J, et al. Identification of novel markers in rheumatoid arthritis through integrated analysis of DNA methylation and microRNA expression. 2013;41:6-16.
128. UniProt. O75716 (STK16_HUMAN) [Available from: https://www.uniprot.org/uniprot/O75716.
129. Liu J, Yang X, Li B, Wang J, Wang W, Liu J, et al. STK16 regulates actin dynamics to control Golgi organization and cell cycle. Scientific reports. 2017;7:44607.
130. Ligos JM, Gerwin N, Fernandez P, Gutierrez-Ramos JC, Bernad A. Cloning, expression analysis, and functional characterization of PKL12, a member of a new subfamily of ser/thr kinases. Biochemical and biophysical research communications. 1998;249(2):380-4.
131. Ohta S, Takeuchi M, Deguchi M, Tsuji T, Gahara Y, Nagata K. A novel transcriptional factor with Ser/Thr kinase activity involved in the transforming growth factor (TGF)-beta signalling pathway. The Biochemical journal. 2000;350 Pt 2:395-404.
132. UniProt. P20132 (SDHL_HUMAN) [Available from: https://www.uniprot.org/uniprot/P20132.
133. Genome_browser. Human Gene SDS (uc001tvg.3) Description and Page Index [Available from: https://genome.ucsc.edu/cgi-bin/hgGene?hgg_gene=uc001tvg.3&hgg_prot=uc001tvg.3&hgg_chrom=chr12&hgg_start=11383025

0&hgg_end=113841692&hgg_type=knownGene&db=hg19&hgsid=725980499_NnWpwwGBFKMw8dDHhV1tSwMAmun5.
134. Kanamoto R, Su Y, Pitot HC. Effects of glucose, insulin, and cAMP on transcription of the serine dehydratase gene in rat liver. Archives of biochemistry and biophysics. 1991;288(2):562-6.
135. Duh EJ, Sun JK, Stitt AW. Diabetic retinopathy: current understanding, mechanisms, and treatment strategies. JCI insight. 2017;2(14).
136. UniProt. Q8NA58 (PNDC1_HUMAN) [Available from: http://www.uniprot.org/uniprot/Q8NA58.
137. Anastasakis D, Skeparnias I, Shaukat A-N, Grafanaki K, Kanellou A, Taraviras S, et al. Mammalian PNLDC1 is a novel poly(A) specific exonuclease with discrete expression during early development. Nucleic Acids Research. 2016;44(18):8908-20.
138. Bouwkamp CG, Kievit AJ, Olgiati S, Breedveld GJ, Coesmans M, Bonifati V, et al. A balanced translocation disrupting BCL2L10 and PNLDC1 segregates with affective psychosis. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2017;174(3):214-9.
139. Yoshida T, Kato K, Yokoi K, Oguri M, Watanabe S, Metoki N, et al. Association of gene polymorphisms with chronic kidney disease in Japanese individuals. International journal of molecular medicine. 2009;24(4):539-47.
140. Schrijvers BF, De Vriese AS, Flyvbjerg A. From Hyperglycemia to Diabetic Kidney Disease: The Role of Metabolic, Hemodynamic, Intracellular Factors and Growth Factors/Cytokines. Endocrine Reviews. 2004;25(6):971-1010.
141. Levey AS, Coresh JJTl. Chronic kidney disease. 2012;379(9811):165-80.
142. Edwards JR, Prozialeck WCJT, pharmacology a. Cadmium, diabetes and chronic kidney disease. 2009;238(3):289-93.
143. Thomas MC, Cooper ME, Zimmet P. Changing epidemiology of type 2 diabetes mellitus and associated chronic kidney disease. Nature Reviews Nephrology. 2015;12:73.
144. Wang Y, Wang B, Yan S, Shen F, Cao H, Fan J, et al. Type 2 diabetes and gender differences in liver cancer by considering different confounding factors: a meta-analysis of cohort studies. 2016;26(11):764-72.
145. Exalto L, Whitmer R, Kappele L, Biessels GJEg. An update on type 2 diabetes, vascular dementia and Alzheimer's disease. 2012;47(11):858-64.
146. Kirpichnikov D, Sowers JRJTie, metabolism. Diabetes mellitus and diabetes-associated vascular disease. 2001;12(5):225-30.
147. Vagelatos NT, Eslick GD. Type 2 Diabetes as a Risk Factor for Alzheimer's Disease: The Confounders, Interactions, and Neuropathology Associated With This Relationship. Epidemiologic Reviews. 2013;35(1):152-60.
148. Xia YY, Ding YB, Liu XQ, Chen XM, Cheng SQ, Li LB, et al. Racial/ethnic disparities in human DNA methylation. Biochimica et biophysica acta. 2014;1846(1):258-62.
149. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clinical epigenetics. 2014;6(1):4.
150. Sommer C, Sletner L, Jenum AK, Morkrid K, Andersen LF, Birkeland KI, et al. Ethnic differences in maternal dietary patterns are largely explained by socio-economic score and integration score: a population-based study. Food & nutrition research. 2013;57.
151. Berntsen S, Richardsen KR, Morkrid K, Sletner L, Birkeland KI, Jenum AK. Objectively recorded physical activity in early pregnancy: a multiethnic population-based study. Scandinavian journal of medicine & science in sports. 2014;24(3):594-601.
152. Zaghlool SB, Al-Shafai M, Al Muftah WA, Kumar P, Gieger C, Waldenberger M, et al. Mendelian inheritance of trimodal CpG methylation sites suggests distal cis-acting genetic effects. Clinical epigenetics. 2016;8:124.

153. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging. 2016;8(9):1844-65.

7 Appendix

Appendix I: PyroMark results

Table 11: The PyroMark results + β values from EPIC BeadChip for cg08098128

cg08098128													
Sample	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**	**Run 7**	**Min value**	**Max value**	**Max - min**	**mean**	**β value *100**	**Abs (β value - mean)**
DNA#1	3,50	4,87	3,37	3,63				3,37	4,87	1,50	3,84	2,63	1,21
DNA#2	3,92	3,66						3,66	3,92	0,26	3,79	2,25	1,54
DNA#3	3,65	3,77						3,65	3,77	0,12	3,71	3,13	0,58
DNA#4	2,64	3,72	3,47					2,64	3,72	1,08	3,28	2,18	1,10
DNA#5	0,00	3,22	0,00	0,00	0,00			0,00	3,22	3,22	0,64	2,12	1,47
DNA#6	3,67	2,75						2,75	3,67	0,92	3,21	2,69	0,52
DNA#7	3,32	3,09						3,09	3,32	0,23	3,21	3,01	0,20
DNA#8	4,15	2,91	3,95	2,92				2,91	4,15	1,24	3,48	2,16	1,32
DNA#9	3,81	2,78	2,28	2,82				2,28	3,81	1,53	2,92	3,76	0,83
DNA#10	3,35	2,73						2,73	3,35	0,62	3,04	1,99	1,05
DNA#11	4,55	2,62	3,76	3,92				2,62	4,55	1,93	3,71	2,80	0,91
DNA#12	3,80	5,70	3,01	4,28	2,00	2,34	2,79	2,00	5,70	3,70	3,42	2,99	0,43
DNA#13	3,74	2,80						2,80	3,74	0,94	3,27	2,08	1,19
DNA#14	3,49	3,03						3,03	3,49	0,46	3,26	2,47	0,79
DNA#15	3,36	2,33	2,62	3,32				2,33	3,36	1,03	2,91	2,61	0,30
DNA#16	2,84	1,99	4,87					1,99	4,87	2,88	3,23	2,95	0,29
DNA#17	3,52	3,77						3,52	3,77	0,25	3,65	2,73	0,92
DNA#18	3,99	3,14						3,14	3,99	0,85	3,57	3,34	0,22
DNA#19	2,54	4,38	4,43	3,40	3,51			2,54	4,43	1,89	3,65	2,72	0,93
DNA#20	3,46	2,65						2,65	3,46	0,81	3,06	2,28	0,77
DNA#21	4,67	4,28						4,28	4,67	0,39	4,48	2,84	1,63
DNA#22	2,89	3,22	2,10					2,10	3,22	1,12	2,74	2,20	0,54
DNA#23	4,09	2,68	3,54	2,90				2,68	4,09	1,41	3,30	3,22	0,08
DNA#24	3,90	4,33						3,90	4,33	0,43	4,12	2,95	1,17
DNA#25	2,81	2,17						2,17	2,81	0,64	2,49	2,20	0,29
DNA#26	4,77	4,37						4,37	4,77	0,40	4,57	2,77	1,80
DNA#27	4,02	2,87	4,46	3,39				2,87	4,46	1,59	3,69	4,22	0,54
DNA#28	4,19	3,40	4,78					3,40	4,78	1,38	4,12	2,44	1,68
DNA#29	3,08	3,48						3,08	3,48	0,40	3,28	2,76	0,52
DNA#30	5,06	3,38	4,64	1,30	3,41	2,79		1,30	5,06	3,76	3,43	2,89	0,54
The average value:										3,76			0,84

Abs = the absolute value

β value *100 = β value from EPIC multiplied by 100 to get methylation range in percentages.

Table 12: The PyroMark results + β values from EPIC BeadChip for cg14120215

cg14120215												
Sample	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**	**Min value**	**Max value**	**Max - min**	**mean**	**β value *100**	**Abs (β value - mean)**
DNA#1	21,85	22,36	20,95	18,57	20,88		18,57	22,36	3,79	21,85	13,47	7,45
DNA#2	5,50	7,89	5,37	4,82			4,82	7,89	3,07	5,50	12,29	6,39
DNA#3	5,55	5,88					5,55	5,88	0,33	5,55	18,28	12,56
DNA#4	5,18	5,18					5,18	5,18	0,00	5,18	15,53	10,35
DNA#5	6,61	6,49					6,49	6,61	0,12	6,61	14,09	7,54
DNA#6	7,38	7,70					7,38	7,70	0,32	7,38	21,28	13,74
DNA#7	9,69	9,07	6,72	7,18	7,98		6,72	9,69	2,97	9,69	14,1	5,97
DNA#8	17,50	17,99					17,50	17,99	0,49	17,50	20,2	2,45
DNA#9	6,28	7,28					6,28	7,28	1,00	6,28	12,45	5,67
DNA#10	5,38	4,20	3,87	5,63			3,87	5,63	1,76	5,38	14,15	9,38
DNA#11	7,16	6,86					6,86	7,16	0,30	7,16	15,36	8,35
DNA#12	7,07	7,07					7,07	7,07	0,00	7,07	22,28	15,21
DNA#13	7,77	8,03					7,77	8,03	0,26	7,77	19,92	12,02
DNA#14	20,22	21,19					20,22	21,19	0,97	20,22	19,05	1,66
DNA#15	16,00	18,51	18,85	15,31	15,41	15,83	15,31	18,85	3,54	16,00	25,93	9,28
DNA#16	3,32	3,89	4,64	6,28			3,32	6,28	2,96	3,32	12,21	7,68
DNA#17	6,46	5,38	7,72	5,44			5,38	7,72	2,34	6,46	12,15	5,9
DNA#18	6,38	5,79	7,69	6,54			5,79	7,69	1,90	6,38	42,13	35,53
DNA#19	5,95	6,20					5,95	6,20	0,25	5,95	12,97	6,89
DNA#20	4,13	2,74	5,32	3,72			2,74	5,32	2,58	4,13	13,25	9,27
DNA#21	5,78	7,37	6,71	6,41			5,78	7,37	1,59	5,78	12,28	5,71
DNA#22	4,58	4,51	5,19	4,75			4,51	5,19	0,68	4,58	14,62	9,86
DNA#23	4,37	5,67	6,21	5,48			4,37	6,21	1,84	4,37	8,26	2,83
DNA#24	4,09	5,40	4,10	3,82			3,82	5,40	1,58	4,09	11,06	6,71
DNA#25	17,81	19,13	22,28	19,85			17,81	22,28	4,47	17,81	11,92	7,85
DNA#26	6,78	6,43					6,43	6,78	0,35	6,78	11,19	4,58
DNA#27	7,64	6,13	5,96	5,31			5,31	7,64	2,33	7,64	14,15	7,89
DNA#28	3,48	3,45					3,45	3,48	0,03	3,48	12,67	9,2
DNA#29	7,06	7,37					7,06	7,37	0,31	7,06	10,58	3,36
DNA#30	5,09	6,02					5,09	6,02	0,93	5,09	15,22	9,66
The average value:									1,44			8,70

Abs = the absolute value

β value *100 = β value from EPIC multiplied by 100 to get methylation range in percentages.

Table 13: The PyroMark results + β values from EPIC BeadChip for cg17148978

cg17148978													
Sample	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**	**Run 7**	**Min value**	**Max value**	**Max - min**	**mean**	**β value *100**	**Abs (β value - mean)**
DNA#1	2,67	0,97	2,28	2,19	0,76			0,76	2,67	1,91	1,77	1,99	0,22
DNA#2	2,77	2,05	2,75	2,31				2,05	2,77	0,72	2,47	2,43	0,04
DNA#3	3,20	2,41	1,71	2,57				1,71	3,20	1,49	2,47	2,56	0,09
DNA#4	2,23	2,09						2,09	2,23	0,14	2,16	2,60	0,44
DNA#5	2,60	1,64						1,64	2,60	0,96	2,12	2,31	0,19
DNA#6	0,00	1,48	1,19	1,30				0,00	1,48	1,48	0,99	3,47	2,47
DNA#7	0,00	2,14	1,80	0,00				0,00	2,14	2,14	0,99	2,57	1,58
DNA#8	1,63	1,65						1,63	1,65	0,02	1,64	2,21	0,57
DNA#9	0,00	1,71	1,01	1,15				0,00	1,71	1,71	0,97	2,82	1,85
DNA#10	1,48	2,29						1,48	2,29	0,81	1,89	2,50	0,62
DNA#11	0,00	2,84	1,51	1,67				0,00	2,84	2,84	1,51	2,54	1,04
DNA#12	1,34	1,83						1,34	1,83	0,49	1,59	1,99	0,41
DNA#13	2,88	1,26	1,45	1,21				1,21	2,88	1,67	1,70	2,07	0,37
DNA#14	2,97	2,35						2,35	2,97	0,62	2,66	2,18	0,48
DNA#15	2,90	2,45						2,45	2,90	0,45	2,68	2,69	0,01
DNA#16	1,42	2,25						1,42	2,25	0,83	1,84	2,09	0,25
DNA#17	1,63	2,41						1,63	2,41	0,78	2,02	2,50	0,48
DNA#18	1,95	1,88						1,88	1,95	0,07	1,92	2,91	0,99
DNA#19	1,42	1,21						1,21	1,42	0,21	1,32	2,69	1,38
DNA#20	1,91	1,92						1,91	1,92	0,01	1,92	2,42	0,51
DNA#21	2,07	1,80						1,80	2,07	0,27	1,94	2,24	0,30
DNA#22	1,77	2,46						1,77	2,46	0,69	2,12	3,20	1,09
DNA#23	2,15	1,55						1,55	2,15	0,60	1,85	2,11	0,26
DNA#24	1,86	2,11						1,86	2,11	0,25	1,99	2,45	0,46
DNA#25	1,73	2,35						1,73	2,35	0,62	2,04	2,82	0,78
DNA#26	1,89	1,72	1,53					1,53	1,89	0,36	1,71	3,73	2,02
DNA#27	1,44	1,63	1,11	1,47	1,07	0,83	1,61	0,83	1,63	0,80	1,31	2,73	1,42
DNA#28	1,47	2,13	1,18					1,18	2,13	0,95	1,59	2,63	1,04
DNA#29	1,48	2,09	1,17					1,17	2,09	0,92	1,58	2,63	1,05
DNA#30	1,09	2,36	1,20	1,88	1,94			1,09	2,36	1,27	1,69	2,66	0,96
The average value:										0,87			0,78

Abs = the absolute value

β value *100 = β value from EPIC multiplied by 100 to get methylation range in percentages.

Appendix II: R-codes used in the master project

TableOne

```
library(foreign)
file.choose()
phenotypes <-
as.data.frame(read.spss("N:\\data\\durable\\Epigenetics\\Phenotype\\SPSSfiler\\IID_phenotypes_EWAS.20.11.18.sav"))
#TableOne
library(tableone)
CreateTableOne(data = phenotypes)
dput(names(phenotypes))
#Vector of variables to summarize
myVars <- c("age", "parity", "b1_sosio_mor_sivil", "sosio_mor_utdn", "røyk_cat",
"b1_inklusjonsuke", "BMI2", "GDM99", "GDM13", "mor_tidligereGDM", "famdiabetes", "gluk0_2",
"gluk2_2", "HOMAIR2", "HOMAB2", "Insulin2", "Cpeptid2", "b2_m_fett_totalt_C", "sObR1")
#Vector of categorical variables that need transformation
catVars <- c("b1_sosio_mor_sivil", "parity", "sosio_mor_utdn", "røyk_cat", "GDM99", "GDM13",
"famdiabetes", "mor_tidligereGDM")
#Create a TableOne object
tab2 <- CreateTableOne(vars = myVars, data = phenotypes, factorVars = catVars)
#Multiple group summary
tab3 <- CreateTableOne(vars = myVars, strata = "etn", data = phenotypes,factorVars = catVars)
tab3
```

Assosiasjon Analysis

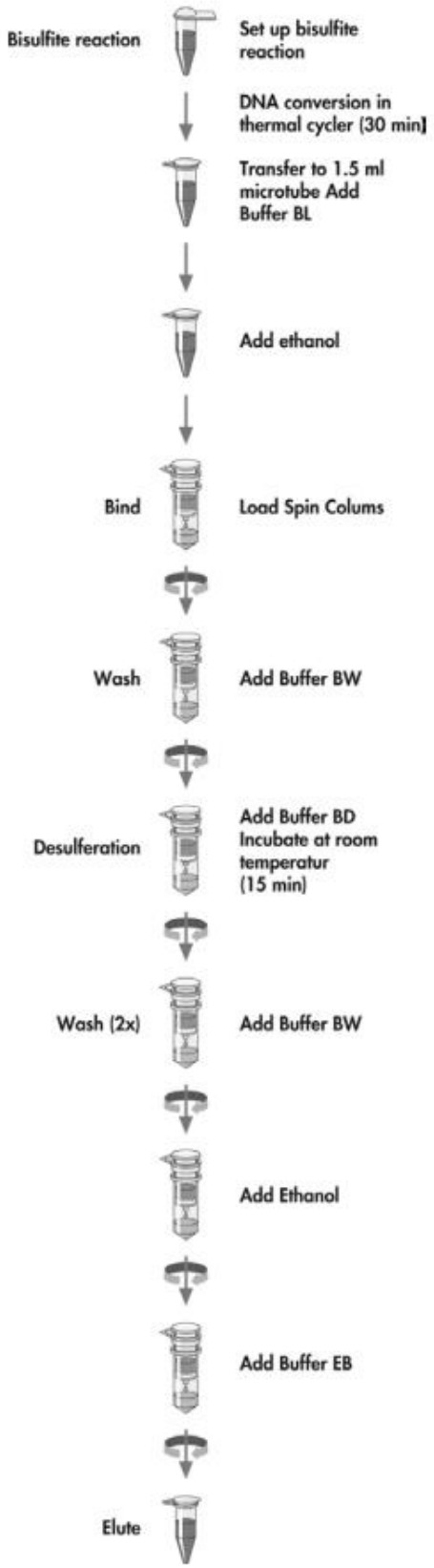
```
# 1.Gjøre klar alle pakkene
library(foreign)
library(tableone)
library(gtools)
# 2. Laster ned filer
pheno <-
as.data.frame(read.spss("N:\\data\\durable\\Epigenetics\\Phenotype\\SPSSfiler\\IID_phenotypes_EWAS.20.11.18.sav"))
#"Nadia_Cpg" laste ned i vindu til høyre "Import Dataset", hvor "coma" endres til "Tab"
#3. Sette sammen datasett
cpgpheno <- merge(pheno, Nadia_Cpg, by='IID')
#4. Sortere etter etnisitet
eu <- cpgpheno[which(cpgpheno$Europe_dum ==1),]
sa <- cpgpheno[which(cpgpheno$SA_dum ==1),]
#5. ta med alle variablene i tabellen
myVars <- c("age", "parity", "b1_sosio_mor_sivil", "sosio_mor_utdn", "røyk_cat",
"b1_inklusjonsuke", "BMI1", "BMI2", "GDM99", "GDM13", "mor_tidligereGDM", "famdiabetes",
"gluk0_1", "gluk0_2", "gluk2_2", "HOMAIR1", "HOMAIR2", "HOMAB1", "HOMAB2", "Insulin1",
"Insulin2", "Cpeptid1", "Cpeptid2", "b1_m_fett_totalt_C", "b2_m_fett_totalt_C", "sObR1", "Ob1",
"Ob2")
catVars <- c("b1_sosio_mor_sivil", "parity", "sosio_mor_utdn", "røyk_cat", "GDM99", "GDM13",
"famdiabetes", "mor_tidligereGDM")
Nonnormal <- c("Insulin1", "Insulin2")
#Lage tabeller med tertiler
satab1<-quantcut(eu$cg19327414, q=3, na.rm = TRUE)
```

```
m1 <- matrix(satab1, ncol=1, byrow=TRUE)
d1 <- as.data.frame(m1, stringsAsFactors=FALSE)
d1[] <- lapply(d1, type.convert)
d1 <- as.data.frame(m1, stringsAsFactors=FALSE)
satert1<-c(d1,eu)
satert2<-data.frame(satert1)
satable1<-CreateTableOne(vars= myVars, strata = "V1", data=satert2, factorVars = catVars)
print(satable1, nonnormal = Nonnormal, quote = TRUE, noSpaces = TRUE)
#Gjenta for hver CpG site bade for EU og SA
```

EPIC – BSP correlation analysis

```
# Legge til a1 EWAS verdier, til b1 BSP verdier
assay_1 <- data.frame(a1,b1)
x_assay1 <- assay_1$a1
y_assay1 <- assay_1$b1
#Gjør det same for Assay 3, Assay 4 og alle assayene samlet)
range <- c(0,0.44)
xrange = range(range)
yrange = range(0:25)
plot(xrange, yrange, main="", xlab="EWAS", ylab="BSP", type = "n")
points(x_assay1, y_assay1, pch=20, cex= 4, col="seagreen2")
points(x_assay3, y_assay3, pch=20, cex= 4, col="skyblue2")
points(x_assay4, y_assay4, pch=20, cex= 4, col="palevioletred2")
points(x_assay, y_assay,cex= 3, pch=1)
abline(lm(y_assay ~ x_assay, data = assay_all), col="midnightblue")
rsq <- function(x, y) summary(lm(y~x))$r.squared
rsq(a,b)
```

Appendix III: DNA bisulfite conversion workflow chart



Appendix IX: MasterMix for PCR of bisulfite converted DNA samples

PyroMark PCR MM 2x	12.5 µl
Coralload concentrate 10x	2.5 µl
Primer (IGFR 10x både fwd + rev)	2.5 µl
RNAse free water	6.5 µl
Template (10 ng/ul)	1.0 µl
Total volume:	25 µl

Appendix X: Agarose gel electrophoresis

Agarose powder	1.5 g
1xTAE buffer	100 ml
Gel red	5 µl