Genome wide methylation analysis and obesity related traits

EMIL NILSSON
Abstract


The most studied form of epigenetics is DNA methylation and several studies have investigated the link between the methylome and body weight. In paper I we analyzed the methylation profile of whole blood in 46 subjects measured with Illumina 27K chip. We provide evidence that obesity influences age driven epigenetic changes. These identified markers may prove to be valuable biomarkers for the understanding of the molecular basis of aging, obesity and associated diseases. In paper II we studied the effect of bariatric surgery, and subsequent weight loss, on methylation and relating this to normal weight controls. In paper II we found 115 promoters had altered methylation after surgery. Among these promoters, an enrichment for genes involved in metabolic processes was found (n=36, p<0.05). In addition, these 51 promoters was more similar after surgery to that of normal-weight controls, than it had been at baseline (p<0.0001). One of the major comorbidities of severe obesity is obstructive sleep apnea and lack of sleep is highly correlated with obesity. Paper III shows how acute sleep deprivation increases portion size and affects food choice in 16 young men. In paper VI, whole genome DNA methylation profiles of whole blood was assessed following both conditions by the Illumina 450K methylation in the same trial as in paper III. This paper shows how sleep deprivation affects DNA methylation profiles of whole blood in a manner both dependent and independent on monocyte subpopulations. Hypothesis free genome wide analysis revealed differential methylation in ING5, a gene previously known to be differentially expressed in sleep deprivation.

Emil Nilsson, Department of Neuroscience, Box 593, Uppsala University, SE-75124 Uppsala, Sweden.

© Emil Nilsson 2015

ISSN 1651-6206
ISBN 978-91-554-9236-6
urn:nbn:se:uu:diva-248685 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-248685)
"Wisdom comes from experience. Experience is often a result of lack of wisdom”
-Terry Pratchett
This thesis is based on the following papers, which are referred in the text by their roman numerals.

   #Equal contribution


IV. Emil K. Nilsson, Adrian Boström, Jessica Mwinyi, Helgi B. Schiöth, Sleep deprivation affects genome wide DNA methylation profiles and RNA expression. (Manuscript)
**Additional publications**

## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASMN</td>
<td>All Sample Mean Normalization</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>BMIQ</td>
<td>Beta MIxture Quartile dilation</td>
</tr>
<tr>
<td>CRP</td>
<td>C Reactive Protein</td>
</tr>
<tr>
<td>DMR</td>
<td>Differentially Methylated Region</td>
</tr>
<tr>
<td>FDR</td>
<td>False Discovery Rate</td>
</tr>
<tr>
<td>IMA</td>
<td>Illumina Methylation Analyzer</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>LGDB</td>
<td>Latvian Genome Data Base</td>
</tr>
<tr>
<td>MUFA</td>
<td>MonoUnsaturated Fatty Acids</td>
</tr>
<tr>
<td>NHANES</td>
<td>National Health And Nutrition Examination Survey</td>
</tr>
<tr>
<td>OSA</td>
<td>Obstructive Sleep Apnea</td>
</tr>
<tr>
<td>PUFA</td>
<td>PolyUnsaturated Fatty Acids</td>
</tr>
<tr>
<td>REM</td>
<td>Rapid Eye Movement</td>
</tr>
<tr>
<td>RYGB</td>
<td>Roux-en-Y Gastric Bypass surgery</td>
</tr>
<tr>
<td>SDB</td>
<td>Sleep Disordered Breathing</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>SWS</td>
<td>Slow Wave Sleep</td>
</tr>
<tr>
<td>TSD</td>
<td>Total Sleep Deprivation</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual Analog Scale</td>
</tr>
</tbody>
</table>
Introduction

Obesity and its associated co-disorders are some of the largest global health problems and are together with the increasing elderly population in many western countries a major challenge for the healthcare system. The risk for several age related diseases such as cancer, type-2 diabetes and neurodegenerative diseases are increased by obesity and overweight (1-3). Hence, it is of utmost importance to understand the mechanisms that underlie these connections and the development of obesity and age related diseases. Recent progress in genetic research has revealed a growing number of genetic variants that predispose carriers to age related diseases, in particular cancers (4). However, the genetic loci associated with obesity have only been able to explain a small fraction of the variation in BMI and other factors are likely more important. The dramatic increase in the incidence of obesity and diabetes seems to have developed over the same time as the progressive decrease in self-reported sleep duration (5). In modern day society, many sleep only 5-6h per night and the trend is continuing (6). It seems restricted sleep is a good predictor of obesity and longitudinal analyses of the 1982-1984, 1987 and 1992 National Health And Nutrition Examination Survey (NHANES) indicate an odds ratio of 2-2.5 of being obese (BMI>30) with 1-4h of sleep (7). An average increase of 2 BMI units was found in subjects with 1-4h sleep compared to 7h sleep. The longitudinal data may be skewed since severely obese subjects often report obstructive sleep apnea (OSA), or sleep disordered breathing (SDB). This may result in a feedforward cascade of negative events generated by sleep loss and hypoxia which likely exuberates metabolic disturbances. Even if data are skewed by this effect, it has been shown that sleep loss contribute to insulin resistance and type 2 diabetes either directly or indirectly by having a deleterious effect on the components of glucose metabolism and appetite regulation (8). Human sleep is normally restricted to a 7-9h period and glucose levels must be maintained during this extended fast. In healthy subjects this is the case but wakeful subjects in a resting position will see their glucose levels fall by 0.04-0.1 mM/h (9). Apart from glucose regulation, sleep disturbances likely leads to obesity by increasing appetite (10), and it has been shown that chronic sleep deprivation leads to marked hyperphagia in rats (11). One hormone that is important for appetite regulation is leptin which is known to increase energy intake. Interestingly, sleep deprivation under continuous nutrition showed a persistence of sleep related leptin decrease indicating that leptin is not only governed by
energy levels (12). This increase in leptin is accompanied with an increase in ghrelin which is also indicative of an impaired glucose metabolism. A systematic review of the links between sleep deprivation and obesity highlights several potential causal mechanisms. Sleep deprivation leads to increased hunger, altered thermoregulation, and increased fatigue, all of which results in increased caloric intake and reduced energy expenditure (13). In the last few years epigenetic alterations have been given an increasing amount of attention as important factors in disease (14) and data suggests that circadian epigenetic patterns are important for rest-activity rhythms, thermogenesis, and satiety (15). In contrast to genetic variations, the epigenetic profile is dynamic and varies with both intrinsic and extrinsic factors throughout lifetime.

Epigenetics and obesity
An important example when discussing epigenetics and obesity is the so called “Dutch hunger syndrome”. The condition emerged as a direct result of prenatal starvation among pregnant Dutch mothers who suffered famine during the winter of 1944-1945 (16). Offspring from these mothers suffered from both cardiovascular and metabolic disease and it was later shown that this was at least in part mediated by epigenetic mechanisms, specifically differential methylation of the insulin-like growth factor II \( IGF2 \) locus (17).

Another recent cohort analysis investigates the correlation between DNA methylation and body mass index (BMI) found three loci related to the gene coding for Hypoxia-inducible factor 3 alpha \( (HIF3A) \) (18). Other examples of links between epigenetics and obesity are the fact that numerous transcription factors and histone modifiers have shown to prevent diet induced obesity (19, 20).

Epigenetics and sleep
Studies have been done showing that epigenetic regulation is present in key genes involved in the sleep-wake cycle, e.g. the circadian clock. The circadian clock instructs 24h rhythmicity on gene expression in essentially all cells and DNA methylation in \( CLOCK \) and \( BMAL1 \) genes, which are crucial for maintaining circadian rhythmicity, are modulated by obesity (21). One of the most consistent degrees of rhythmicity, phasing and amplitude of gene expression is found in the 5-methyltetrahydrofolate homocysteine methyltransferase \( (MTR) \). This gene is involved in DNA methylation which is a core epigenetic process in humans (22). Other epigenetic processes such as RNA editing and histone modifications have been shown do display diurnal patterns but the most studied form of epigenetics is DNA methylation (23). In
mice, the period circadian clock gene \((PER1)\) is regulated by DNA methylation (24) and DNA methylation of the frequency gene in neurospora is involved in setting the proper phase of the circadian clock (25). Some studies suggest that monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids could play a role in DNA methylation and epigenetic regulation of the circadian system (26). These observations suggest roles for DNA methylation in circadian clocks and that clock gene methylation profiles are involved in many diseases such as obesity, metabolic syndrome and cancer (21, 27).

Epigenetics of the cell

Cells employ epigenetic mechanisms mainly for three purposes; cellular memory, differentiation, and adaptation. Cellular memory includes mechanisms aimed at maintaining cellular identity. These mechanisms become important after a differentiation event where cell identity is determined. Often, a single differentiation event is enough to establish a cellular memory that lasts throughout the lifespan of all daughter cells. Differentiation is important in embryonic development and lately, the epigenetic adaptation in response to environmental stimuli has received increasing attention from the research community (28). The maintaining, differentiation, and adaptation of complex gene expression patterns are governed by molecular mechanisms working in all stages of transcription. DNA methylation and chromatin modifications modulate transcription initiation, microRNA have the ability to degrade mRNA transcripts and circulating transcription factors that can be inherited from mother to daughter cell provide a molecular memory that does not require any transcription in the daughter cell in order to function.

DNA methylation

One of the most studied epigenetic mechanisms is the methylation of cytosine residues, which is maintained and controlled by different DNA-methyltransferases (DNMTs). Cytosines across the genome tend to be methylated (29), but in Cytosine phosphate Guanine (CpG) rich regions in proximity of genes the methylation is dynamic and functions as a gene specific regulatory mechanism of transcription (30). Such cytosine enriched regions are called CpG islands and a higher methylation in this type of region is often associated with a reduced expression of the nearby gene, due to chromatin rearrangement, inhibition of transcription activators and/or recruitment of transcription repressors (31-33). Hence, DNA methylation provides a regulatory mechanism of gene transcription and is essential for cell fate, differentiation and tissue integrity. The methylation status of monozygotic twins diverge with age, which demonstrates that DNA methylation is susceptible to
environmental factors (34). This strengthens the notion that the methylome is an adaptive entity capable of changing an individual's gene expression pattern due to environmental factors. In fact, it has been demonstrated that factors such as diet and nutrient intake affect the methylation status as well as conditions such as inflammation, oxidative stress and hypoxia (32). Several studies have reported genomic regions where the methylation level is differentiated in obese individuals and varies with body-weight (35-38). In contrast, methylation levels are also associated with genetic variations and can thus be governed, at least in part, by genetic factors (39). With age the global methylation level of the genome (including non-CpG regions) is decreased, which leads to a global hypomethylation (40, 41). This may be caused by a lower expression of DNMT1 with age, which would lead to a slower de novo and maintenance of methylation. In contrast, several CpG islands in promoter regions are hypermethylated during aging (42-44). While this demonstrates that the genome can be locally hypermethylated during aging, it is unclear if this is a general process for promoter associated CpG islands. These age associated methylation changes could be important in age driven diseases and has received particular attention for its potential important role in cancer (45).

Measuring DNA methylation

DNA methylation occurs as a covalent bonding of a methyl group to the 5’ position in the pyrimidine ring of cysteine residues. In recent years we have seen increased availability of methods for detection and quantification of DNA methylation. Early methods able to quantify the global average methylation were often based on high performance liquid chromatography (HPLC) but the bulk of methods in use today rely on sodium bisulfite conversion (46) followed by single loci or genome wide quantification steps. Sodium bisulfite converts unmethylated cytosines to uracil making detection of methylation state possible by the vast array of sequence based methods.

![Figure 1: Bisulfite conversion of cytosines to uracil: Cytosines can have a double bond between the 4’ and 5’ carbon in which case a sulfate group can attack and subsequent deamination and desulfonation results in uracil. If there is a methyl group on the cytosine the initial sulfonation step is not possible due to there not being a reactive electron pair.](image-url)
The first method to employ bisulfite conversion to measure methylation in a locus specific manner was methylation specific PCR (MSP or MS-PCR) (47). In this method primers are designed to amplify the product of either methylated or unmethylated DNA. The level of methylation can then be assessed by comparing the efficiency of one primer pair over the other. It is also possible to design methylation unspecific primers that amplify the DNA regardless of methylation state. The product is then available for sequencing for quantitative measurement. In contrast to bisulfite sequencing, MSP is not quantitative and variations of MSP such as MethylLight (48) or quantitative analysis of methylated alleles (QUAMA) (49) have been developed. Recent advances in microarray and sequencing technology have made available genome wide analysis of DNA methylation.

Genome wide methylation analysis by array based methods

Perhaps the most widely used methods for whole genome methylation profiling is the Illumina BeadChip. The Illumina 27K BeadChip is designed by the manufacturer to preferentially target CpG sites in proximity to the promoter of 14,475 genes of the consensus coding sequences (CCDS) and known cancer genes as well as the promoter of 110 miRNA promoters. Hence, the array is designed to study CpG sites in proximity to genes and not the methylation of intergenic cytosines or repeat regions. The Illumina 450K is a major upgrade to the earlier 27K chip and investigates over 15 times the amount of CpG sites (>485000) compared to its predecessor. This addition in probes means that the Illumina 450K is not only targeted towards CpG islands in promoter regions of genes. Apart from low and intermediate class islands, regions in intergenic and downstream regions of genes are also investigated (50). In fact, the 450K chip accomplishes this by employing both type I and type II probes while the 27K only uses type I probes. Type I probes use two bead types or probes per methylation loci, one for each of the methylated and unmethylated states. After successful hybridization of one of the probes, the methylation state is determined by single-base primer extension.

Figure 2: The effect of bisulfite conversion on DNA sequence. The DNA sequence is converted so that unmethylated cytosines are transformed to uracil. The target CpG site is then identified by single base-pair extension.
Type II probes are universal and hybridize regardless of methylation state, the single-base extension step then occurs at the CpG site and the type of incorporated nucleotide determines the methylation signal. Both chip designs allow for 12 samples to be processed on the same chip. In paper II and IV, where we employ a within subject design, the samples from the same individual was hybridized to the same physical chip in order to avoid inter-chip biases.
**Aims**

The aim of this thesis was to investigate the connection between obesity, epigenetics and sleep. In the last years, epigenetic phenomena and intergenerational effects have been considered potential contributors to the obesity epidemic. Our growing recognition of the overall importance of epigenetics in the circadian clock and in associated diseases has major implications. This includes the potential to define the molecular basis for the poorly understood relationships that exists between circadian processes and neural developmental and neurodegenerative disorders, metabolic diseases, cancer and aging. Much recent research has focused on the crosstalk between the circadian clock and metabolism. Crosstalk between circadian rhythm and metabolism is essential for maintaining homeostasis and prevent metabolic disorders. Moreover, food availability and energy intake participates in this crosstalk, regulating the circadian clock as well as metabolic pathways. Paper I represents a first step in investigating the relationship between DNA methylation status, obesity in conjunction with age. In II and IV we conduct further hypothesis free exploratory studies aimed at identifying new molecular markers and biological pathways that may regulate or be regulated by obesity and sleep. In paper III we aim to investigate the effects of sleep deprivation on food intake, thus further defining the complex dependency between obesity and sleep.
Materials and methods

Ethics statement and subject description relating to paper I

Written confirmed consent was acquired from all participants of the study. The study was approved by the Central Medical Ethics Committee of Latvia. Study samples were acquired from the Latvian Genome Data Base (LGDB), a national biobank of health and genetic information collected for adult residents of Latvia (over 18 years old). Health status of the participants was asserted by health care professionals according to International Classification of Diseases (ICD-10) codes. Information on a familial health status, ethnic and social background, lifestyle and anthropometric measurements were obtained in a questionnaire based interview. We selected 24 obese and 22 lean female adults from a total group of 934 females, with a known FTO rs9939609 genotype, that were recruited to LGDB from 2003 to May 2009. Selection criteria included rs9939609 genotype, Body Mass Index (BMI) (lean <25 kg/m2 and obese ≥30 kg/m2) and health status (participants diagnosed with endocrine diseases and malignant tumors before recruiting to LGDB were not included). Middle age females were selected and a comparable age range of the lean (41-69 years old) and obese females (42-70 years old) were ascertained. The individuals were selected so that both the obese and lean groups were composed of equal proportions of homozygous carriers for the normal and risk allele of the rs9939609 SNP. Heterozygous individuals were excluded. Age weight and BMI details of the participants can be found in Table 1.

Ethics statement and subject description relating to paper II

Eleven obese patients who qualified for bariatric surgery according to international guidelines (51) and were willing to undergo a RYGB procedure were included in the study. During the RYGB procedure the largest part of the stomach was transected, and a small gastric pouch of about 20-30 ml was then anastomized to the proximal jejunum with the diameter of the pouch–jejunal anastomosis standardized to be about 12 mm. In addition, the biliopancreatic limb is side to side anastomized to the jejunum 150 cm distal
from the pouch–jejunal anastomosis (Roux-en Y limb length, 150 cm). Patients were recruited into the study at their time of surgery between 20090820 and 20110210 and were followed up 6 months after the surgery. Thus patient follow-up was performed February 2010 for the first patient and August 2011 for the latest patient. The subjects were chosen so that they were as healthy as possible apart from being obese. Written informed consent was obtained from all subjects, and the study was approved by the ethic committee of Canton St. Gallen, Switzerland. Ethical approval for the study was obtained 20080717. A consort flowchart of the selection procedure is available as figure 1. The control group comprising 16 normal-weight men was obtained from the control condition of an as of yet unpublished study (Clinical trial number: NCT01730742). This study was approved by the regional ethic committee in Uppsala (regionala etikprövningsnämnden i Uppsala, www.epn.se), Sweden and written, informed consent was signed by all participants. For the determination of promoter-specific DNA methylation, whole blood was sampled after an overnight fast in EDTA coated tubes, both at baseline (~1-2 weeks before the surgery) and six months after RYGB. In the control group, whole blood was sampled in the morning after nocturnal sleep (data are not shown). Until assay, all whole blood samples were kept frozen at -80 C. Weight was measured while subjects were wearing light close. Fasting plasma glucose (FPG) levels were measured by routine clinical laboratory analyses.

Ethics statement and subject description relating to paper III and IV

16 healthy normal-weight male subjects (BMI = 23.6 ±0.6, age = 23 ±0.9) were included in the study. Prior to the experiment, all subjects reported a normal sleep/wake rhythm with 7-8h of sleep starting between 22:00-23:30 and ending 06:00-07:30. They were not on shift work, ate breakfast regularly, and were not on any medication. Exclusion criteria included a history of medical disorders or sleep complaints. Subjects underwent a physical examination coupled with routine laboratory testing including C-reactive protein concentration and white blood cell counts. The study is registered with www.ClinicalTrials.gov (NCT01730742) and all participants gave written informed consent before participating in the study. The study itself employed a randomized and balanced within subject design where all subjects participated in two experimental conditions: sleep and total sleep deprivation. The experimental night was preceded by a 28.5 hours long baseline period consisting of regulated and monitored sleep, exercise, and food intake. The subjects arrived at 18:00 (day 0) and received a standardized meal. They were then allowed to sleep for eight hours (22:30-06:30) and lived in the standardized environment until 20:30 the following evening (day 1) when they were
informed about whether or not they would be sleeping the coming night. Polysomnography was performed by use of Embla A10 recorders (Flaga hf, Reykjavik, Iceland) in order to ensure compliance. The polysomnography comprised of electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG). Sleep stages were determined using the criteria previously described by Hobson et. al. (52) by an experienced scorer blinded to the study hypothesis. After the experimental night (day 2) they were subjected to a battery of blood and behavioral tests in a fasted state between 07:30 and 09:00. The test included CRP, white blood cell count, serum and plasma samples, glucose measurement, and a portion size task. Plasma concentrations of total ghrelin were assayed by a commercially available kit (EZGRT-89K, Millipore, Billerica, MA, USA).

DNA isolation and bisulfite conversion

DNA isolation was performed as we have previously described using the phenol chloroform method (53). Bisulfite conversion was performed with the EZ DNA Methylation-Gold™ kit (Zymo research): 500 ng of DNA was subjected to bisulfite treatment including heating to 98°C for 10 min followed by conversion at 64°C for 150 min. The bisulfite conversion converts all (> 99%) unmethylated cytosine to uracil, which gives rise to a DNA sequence that can be defined by its initial methylation status.

Whole genome microarray hybridization in papers I, II, and IV

In paper I, the Illumina Infinium HumanMethylation27 BeadChip array (Illumina) probes 27,578 different CpG sites across the whole genome and has been shown to yield reproducible results in agreement with technologies such as bisulfite sequencing (54). In paper II the degree of DNA methylation was determined by the Illumina 450K methylation chip (Illumina, San Diego, USA). Both methylation chips use the same initial steps. The DNA was whole-genome amplified, enzymatically fragmented, precipitated, re-suspended and after hybridization overnight at 48°C the difference between a C or a T nucleotide was detected by single-base primer extension. The fluorescent detection was done using the Illumina iScan scanner. Paper I and II the ratio between the signal from the C and the sum of the C and T signals, was performed with the GenomeStudio 2009.2 (Illumina) software. In paper IV this process was removed and preprocessing began from the IDAT files produced by the iScanner.
Meal size estimation in paper III

The subjects were asked to choose their ideal portion size of 13 commonly consumed foods, seven of which were meal items (spaghetti Bolognese, penne and tomato sauce, rice curry (chicken tikka masala), boiled potatoes, oven fries, baguettes with garlic and herb butter, and cheese and tomato pizza) and six were snack items (cashew nuts, pretzels, Pringles, banana, KitKat and peanut M&Ms). The process measures satiety expectations and relies on the methodology ‘method of constant stimuli’(55). The experiment was carried out on a laptop before breakfast on day 2 of the trial, both before and after the standardized breakfast. The ideal portion size was selected by using the arrow keys on the keyboard and thereby displaying the desired portion ranging from 83 to 750 kcal. The subjects were also asked to rate their perceived hunger on a 100mm VAS scale both before and after the breakfast.

Microarray preprocessing in paper I, II, and IV

All preprocessing and analysis was done using the statistical software R (www.r-project.org) in conjunction with the limma package from Bioconductor (56, 57). Papers I and II used beta values (value ranging from 0 to 1 reflecting 0% to 100% methylation) obtained from Genomestudio (Illumina, San Diego, USA) as the starting point for preprocessing and analysis. In paper I, Downstream analysis was made using the entire dataset (excluding only high detection p-values). In paper II and IV, preprocessing was more extensive due to the more complicated nature of the Illumina 450K which was described earlier (58). In paper II, beta values were preprocessed using the Illumina Methylation Analyzer (IMA) package for R (59). Probes were excluded if they met any of the following criteria: at least one SNP in probe region matched multiple genomic regions, targeted sex chromosomes or had a detection p-value > 10^-5. In paper II, this filtering was followed by peak correction (60) and quantile normalization. In addition an average on all CpG sites located in the same promoter region was calculated as this analytical approach is proposed to be more meaningful from a biological perspective (61). In total, 16724 promoter regions were used for subsequent analysis in paper II. In paper IV, we employed an even more extensive preprocessing paradigm. GenomeStudio was bypassed in favor of the All sample Mean Normalization (ASMN) approach (62). Probe filtering was similar in paper II and IV with one exception. Paper IV used more recent annotation data to remove SNP related probes as well as ambiguous probes (50). Another difference is that while paper II averaged promoter associated probes for the same gene paper IV employed a broader approach based on more recent results (63). In paper IV, all probes within 2000 bp of the TSS was used in downstream analysis resulting in 167490 eligible probes. These probes went
through quantile normalization and Beta Mixture Quantile dilation (BMIQ) before any analysis was made.

Statistical analysis in paper I
Linear regression of methylation and age was performed using the lmFit function in limma (robust regression, 1000 iterations). A Benjamini-Hochberg p-value < 0.05 was considered significant. For genes with significant probes, all other probes for that same gene were also investigated. A nominal p-value < 0.05 was considered significant for these adjacent probes. Each probe was investigated for potential interaction between age and weight group (obese and lean) using general linear models. Probes that had a significant interaction were also analyzed within the obese and lean group separately to retrieve the group specific effect of aging on the methylation level of the site. Furthermore, the probes with no significant interaction were analyzed without the interaction term to detect the main effect of obesity and age on the methylation level. The influence of the FTO rs9939609 genotype, which lies in intron one and is not likely to have any cis effects, on the methylation pattern was investigated by implementing a linear model controlled for age and weight group. The significant genes were analyzed for enrichment of function using the Consensus database (64, 65) with all the 14,446 genes represented in the Illumina Infinium HumanMethylation27 BeadChip array that passed QC used as a background. Both KEGG (66), the Kyoto Encyclopedia of Genes and Genomes (which catalogues genes based on the biological pathway they are involved in) and level 4 biological process GO terms (67) were used for enrichment analysis.

Statistical analysis in paper II
The initial analysis aimed to examine if RYGB surgery would reduce the genome-wide distance between promoter DNA methylation of obese patients and the control group (comprising normal-weight men). To this aim, the Euclidean distance of all promoters (i.e. 16724) was estimated for both presurgery time point vs. control group and postsurgery time point vs. control group. These distances were then compared by means of a pairwise t test. In the second analysis, presurgery DNA methylation of the above mentioned 16724 promoter regions was compared with that obtained at six month after RYGB (controlled for multiple comparisons, i.e. Bonferroni corrected). All significant promoter hits revealed by this pre/post comparison were then subjected to an additional Euclidean distance analysis. Overall, a p-value < 0.05 was considered significant. Enrichment analysis was performed using the ConsensusPath database (64).
Statistical analysis in paper III

The initial analysis compared the effect of TSD on portion size for the most liked items. The items were chosen by a median split resulting in 7 of the food items, 4 meal items and 3 snack items. A multivariate ANOVA was used to explore the effect TSD and/or breakfast received on perceived hunger ratings, ghrelin concentrations and selected portion size. The pairwise comparison between sleep and TSD used weighted contrasts and was analyzed using SAS (version 9.5; SAS institute Inc.). All data are represented as means ± SEM and overall, a p-value < 0.05 was considered significant.

Statistical analysis in paper IV

A pairwise linear regression for sleep vs sleep deprivation adjusting for the ratio of neutrophils to leukocytes was fitted to each probe. The optimal p-value threshold was determined by maximizing the ratio of observed significant hits over the number of expected false positives. The significant probes were then selected as seeds for a second analysis where the change in methylation was compared to the median change in technical replicates. In this step a Benjamini-Hochberg corrected p-value < 0.05 was considered significant. A nominal p-value of 0.05 was considered significant for adjacent probes. Probes relating to genes previously shown to be differentially expressed in sleep deprivation were chosen for a targeted analysis (68). Additionally, the relation between DNA methylation and mRNA expression for significant probes were assessed using a separate cohort (GeoID: E-GEOD-49065). Enrichment analysis of biological pathways and molecular interactions were performed using Consensus path and GeneMANIA (69) respectively.
Results

Results of genome wide methylation analysis investigating variations with age and obesity

In paper I, we wanted to ascertain whether age, obesity and their interaction could predict the methylation status of specific CpG sites using an array that predominantly targets regions in proximity of promoters. We also sought to replicate previous findings that a genetic variation within the FTO gene is associated with methylation changes (70). A linear model was implemented where the methylation of each site was evaluated as a linear function of obesity, age and their interaction term. In the case of the age associated genes, statistical analysis using ConsensusPathDB was used to identify enriched functional clusters among the differentially methylated genes. Significant correlation between age and methylation levels after Benjamini-Hochberg correction (FDR < 0.01) was observed in 125 probes of which 13 also proved significant under Bonferroni correction. 70 of these 125 probes were annotated to genes that had multiple probes associated to age with a nominal p-value < 0.05. Of all 125 probes, 34 showed reduced methylation with age and 91 were hypermethylated. An enrichment analysis of functional and biological terms revealed the KEGG pathway map04080 “Neuroactive ligand-interceptor interaction” to be enriched in this dataset (FDR<0.01). Nine of the members of this KEGG group (PTGDR, MTRN1A, PRLHR, HTR7, MLNR, GRIA2, GRM1, GLRA1, THRB) were found among our age related sites. We identified an additional 10 regions after Benjamini-Hochberg correction where the methylation levels depended on the interaction between obesity and age and analyzed them separately in the lean and obese group (paper I, Figure 1). In eight (ADCY1, CXADR, KCNS2, LMX1B, FNDC4, NAT8L, AQPEP and FBLIM1) of the ten cases the obese subjects displayed decreased methylation with age when compared to their lean counterparts, whereas the opposite was true for the remaining two sites (RNH1 and NNAT). The gene “Long intergenic non coding RNA 304”-LINC00304 (Illumina ID: cg03819692, position chr16:87753140, located 11 bp from the transcription start site) displayed higher methylation (p-value = 0.0030, adjusted with Benjamini-Hochberg, FDR < 0.001) in the obese individuals compared to the lean, independent of age (paper I, Figure 2a). No other gene was found to be differentially methylated between obese and lean individuals. Furthermore, no differential methylation level could be detected between
the normal and risk allele carriers of the FTO gene. The average methylation level of all probes on the array was calculated. This value was fitted to a linear model and investigated for correlation with age and obesity. A trend for the average genome-wide methylation levels to increase (p-value = 0.10) with age was observed. No differential global methylation level was detected between the obese and lean individuals (p-value > 0.25). It is important to stress that the genome-wide average methylation does not reflect the global methylation, which includes intergenic sites that are underrepresented on the array.

Results of genome wide methylation analysis investigating the effect of bariatric surgery on methylation profile

The mean genome-wide Euclidean distance between promoters of obese patients at six months after RYGB surgery and controls was significantly shorter, as compared to that at baseline (p-value < 0.001). An additional analysis demonstrated that at six months after RYGB surgery, the DNA methylation of 51 promoters was significantly different from corresponding presurgery values (28 were upregulated and 23 were downregulated, p-value < 0.05 for all promoters, Bonferroni corrected). The mean DNA methylation of these 51 promoters was also more similar after surgery to that of controls, as compared to that at baseline (p-value < 0.0001). Importantly, when controlling for the RYGB surgery-induced drop in weight (~24% of respective baseline value) and fasting plasma glucose concentrations (~16% of respective baseline value), the DNA methylation of only one out of 51 promoters (~2%) remained significant at the postsurgery time point. An enrichment analysis of GO-terms biological processes using DAVID functional analysis revealed an enrichment for genes involved in metabolic processes (GO:0008152) (FDR < 0.05).

Results from sleep deprivation experiment

As mentioned, paper III and IV are based on the same sleep deprivation experiment. Sleep characteristics indicate a full night’s sleep in the control condition (total sleep time = 442±6 min, wake = 30±6 min, sleep stage 1 = 5±1 min, sleep stage 2 = 219 ± 11 min, slow-wave sleep = 115±6 min, REM sleep = 103±8 min, sleep efficiency = 92±1%). Plasma concentration of CRP was less than 6 mg/ml indicating an absence of infection and white blood cell counts were normal (5.2±0.7).
Results from portion size, hunger and ghrelin measurements

The portion size was larger after TSD compared to sleep both in the fasted state where there was an increase of 11% (p-value = 0.07) and in the sated state where the increase was 6% (p-value = 0.01). The effect was most prominent when considering the most liked items where there was an increase of 14% (p-value = 0.02) in the fasted state and 12% (p-value < 0.01) in the fasted state. Moreover, in contrast to the portion size pattern before breakfast, subjects selected larger portions of snack but not meal items in their TSD condition, 14% (p-value = 0.02). An increase in portion size was accompanied with an increase in hunger rating after TSD in both before and after breakfast (24%, p-value < 0.01 and 15%, p-value = 0.04 respectively).

As expected, ghrelin concentrations were also higher in the TSD condition, 13% (p-value = 0.04).

Genome wide methylation analysis in sleep deprivation

We discovered 269 probes that were differentially methylated in sleep vs sleep deprivation (Supplementary table S1). Figure 2a (paper IV) depicts the quota between observed and expected number of significant probes correlated to TSD and N/L. A nominal p-value threshold of 10^-2.5 was decided based on this figure. This corresponded to a maximum ratio of 1.22 between observed and expected number of hits. Consensus path (64) enrichment analysis with all analyzed probes as background revealed two pathways to be enriched among the DMR. The enriched pathways were the NOTCH (NOTCH4, RING1, HES1, CCND1, MAML1, FDR<0.05) and WNT (CCND1, FZD8, FRAT1, FZD6, WNT4, FDR<0.05, source: Wikipathways (71)) signaling pathways. The 269 probes were mainly situated in high density CpG regions (see figure 2b, paper IV). The opposite was true for N/L ratio correlated probes. Both TSD and N/L ratio linked probes displayed a different distribution than the background (chi-squared test, p-value < 0.05).

Many probes shared its closest gene with other probes and 119 (44%) of the genes in supplementary table S1 (paper IV) had more than one significant probe. Even if no single probe displayed a correlation with expression large enough to pass Benjamini-Hochberg correction, there was an enrichment of low p-values. 26 of the 237 probes with available expression data showed a correlation between methylation and expression (p-value < 0.05) compared to 11.85 which would be expected by chance (binomial p-value < 0.001). We directed an analysis of methylation-methylation correlations for the 100 high-CpG promoter probes (within 2000 base-pairs upstream of TSS) and this is represented by the correlation plot in figure 3 (paper IV). The probes connected to the enriched pathways (NOTCH and WNT signaling) and are highlighted in this plot. In the NOTCH pathway, three of the probes (RING1, MAML1, HES1) were located in the upstream region of their gene and two (NOTCH4, CCND1) were located downstream. The two downstream genes
correlated well with the methylation of the upstream probes (R = -0.38, sd = 0.06 across all comparisons). Two of the WNT probes are next to each other in the correlation plot and a correlation test between these two and \textit{CCND1}, \textit{WNT4}, and \textit{FZD6}, showed a correlation trend where the p-value for all three comparisons were less than 0.1. This indicated that there is a correlation between the upstream and downstream probes in both the WNT and the NOTCH category. One of the 269 probes was situated 69 base-pairs upstream of the \textit{ING5}-gene in a high density CpG region, a gene that was also shown to be differentially expressed in total sleep deprivation by Möller-Levet et. al. (68). They also reported differential expression of 121 other genes which corresponds to 1631 probes on the Illumina 450K chip. In an identical paired analysis coupled with comparison of technical replicates as used initially these probes were analyzed specifically and 4 methylation probes proved significant under Benjamini-Hochberg correction in this targeted analysis (see figure 4, paper IV).
Conclusions

In paper I we found one loci where there was differential methylation between obese and lean individuals. The novel epigenetic marker is in the proximity of the gene \textit{LINC00304}. The exact function of \textit{LINC00304} is unknown but long intergenic non-coding RNA’s are often associated with transcriptional regulation. There was a considerable age-dependency in the data and we determined that 135 genomic sites are subject to differential methylation during aging and that this process is influenced by weight in a subset of loci. Genes involved in neuroactive, ligand-receptor interaction are overrepresented among the age-related probes (n=125). We identified 10 sites with an interaction effect between obesity and aging, one of which is \textit{LMX1B}, a gene recently identified as an obesity susceptibility loci (24). Intriguingly, the gene is known to be involved in both obesity and age realated diseases (72). The transcription factor \textit{LMX1B}, is involved in the development and maintenance of dopaminergic neurons, and is associated with Parkinson's disease (72). Another age related DMR reported in paper I is situated close to the \textit{NNAT} gene which is associated to obesity (73). The \textit{NNAT} is regulated by leptin and is an example of what is considered an imprinted gene. Our results regarding this gene could therefore reflect a differance in infancy between subjects. Imprinting is a potential confounder in cross sectional epigenetic studies and in paper II and IV we limit this by using paired samples. In paper II we use a similar number of subjects (24 in paper II vs 47 in paper I) but were nevertheless able to identify 51 differentially methylated regions. Moreover, when comparing the genome wide distance between pre and post surgery to controls, we demonstrate that RYGB decreases the genome wide distance between promoter-specific DNA methylation in whole blood. The shortened distance was most pronounced among the 51 DMR's but was also detected on a genome wide level. This indicates that a majority of the methylation changes in the obese patients 6-months after surgery was related to their shared conditions, i.e. RYGB and subsequent weight loss. As one might expect, metabolic processes were enriched among genes with promoter probes whose DNA methylation was significantly changed at 6 month after RYGB surgery. Several of these might account for some metabolic and clinical benefits that are typically observed. For instance, \textit{INCA1}, whose methylation was about 12% lower at 6 month after RYGB than it had been at baseline, may have an anti-cancer effect due to its anti-proliferative properties (74). As discussed previously, one of the factors that may cause
obesity is sleep deprivation induced hyperphagia (11). In paper III we demonstrate that total sleep deprivation (TSD) increases both feelings of hunger and plasma ghrelin levels the following morning. This translated into increased portion size regardless of the type of food offered before breakfast. However, postprandial portion sizes were only larger in the snack category. The fact that sleep deprivation increased portion size in the fasted state irrespective of the type of food offered suggests that overeating following sleep loss may represent a homeostatic compensatory response. This homeostatic response is likely produced to compensate for the energy deficits that result from sleep deprivation (9, 75). Another factor may be the increase of leptin associated with TSD (12) which together with increases in ghrelin plasma concentrations may represent a hedonic response involving appetite regulation. Epigenetic responses to total sleep deprivation has previously been observed for specific genes e.g. CLOCK, PER1 and BMAL1 (21, 23, 27) and in paper IV we demonstrate that TSD induces significant changes in the epigenetic profile in blood, which are partly independent from the ratio of neutrophils over leukocytes (N/L ratio). CpG sites relating to 269 different genes were found to be differentially methylated after TSD. One of the detected CpG sites belongs to the gene ING5 and has previously been shown to be differentially expressed as a result of TSD (68). ING5 is a tumor repressor gene that is acting by acetylating p53 in response to DNA damage (76). Naturally, disruption of ING5 has been linked to several types of cancers, particularly gastric cancers (77, 78). A directed analysis based on previously known differential expression in sleep deprivation revealed additional three genes (CIT, USP46, and FGFR1OP2) where methylation affected by TSD in humans. In mice, the USP46 is responsible for the phenotypes found in the CS mouse strain. CS mice exhibit several distinct phenotypes related to circadian behavioral rhythms and tail mobility, the latter is a phenotype used to measure depression-like behavior in mice (79). SNP analysis in a Japanese population confirmed USP46 as a gene linked to major depressive disorder (MDD) in humans (80). CpG sites within genes of the NOTCH and WNT signaling pathway appeared to be enriched after sleep deprivation. NOTCH signaling has been especially associated to embryonic development and cell fate (81). Disruptions in cell fate determining mechanisms may lead to various cancer forms. The WNT pathway plays a role in cancerogenesis as well. Stem cell cancer growth can be stopped by inhibiting key features of notch and WNT signaling (82). In line with our results, Both NOTCH and WNT signaling pathways have earlier been linked to sleep in different ways. NOTCH signaling modulates sleep homeostasis in Drosophila by regulating the expression of the Bunched gene, which modulates the sensitivity to sleep loss (83). The association between NOTCH and sleep homeostasis was hitherto mainly demonstrated in nerve cells but the DNA methylation profile show a strong overlap between blood and brain (84, 85), which made it reasonable to perform our studies whole blood. The results of these studies indicate a bidirectional association between sleep, the immune system and...
inflammatory markers (86, 87). This is likely triggered by inflammatory cytokines in a manner similar to those found in depressed or alcoholic patients (86). Based on the obtained results we hypothesize that TSD dependent changes in blood cell composition are reflected in N/L dependent shifts in the methylation profile, which are especially located in low-CpG regions. In contrast to that, N/L ratio independent methylation changes induced by TSD were detected high density CpG regions. Several limitations apply in the genome wide methylation experiments. In paper I the size and specificity of the sample is the major concern coupled with the fact that it is a cross-sectional study which is sensitive to imprinting events. Moreover the relatively small changes reported is cause for caution since technical replicated in 5% of the probes display a difference of 13.6% (88) and the biggest change is that of SERHL which changes -12% between the oldest and the youngest subject. In paper II and IV we attempt to alleviate problem with cross-sectional studies by using paired samples. In paper IV we also take monocyte subpopulations into account. All methylation studies and measurements were performed in whole blood and extrapolation of these results on methylation patterns in other tissues should be done with caution. Despite our limitations we are able to provide new evidence of the link between obesity and chronobiology. However, considerable effort is required to unravel the complexity of these epigenetic, hedonic, homeostatic, and environment interactions and to evaluate their potential reversibility.
The aim of this thesis was to investigate the complex nature of the interplay between sleep and obesity. In paper III we present a causal link between sleep deprivation and obesity via increased portion sizes. It is clear from the results in paper I that not only is DNA methylation largely determined based on other factors i.e. age, but these different phenotypes interact with each other to form progressive methylation patterns that were discernible between old and young obese subjects. We were able to define epigenetic marks connected with both obesity and total sleep deprivation in papers II and IV. This means that if there is an epigenetic link between sleep and obesity, the direction and mechanism of it remains to be elucidated. Nevertheless, it is likely that at least some of the epigenetic dysregulation preceding or succeeding obesity can be affected by novel treatment options. In rats for example, pancreatic β-cell deficits and impaired glucose tolerance in small weight offspring is partially ameliorated by improved postnatal lactation (89). Other researchers are investigating the possibility of targeting epigenetic mechanism with small molecule drugs (90). Identifying the best epigenetic targets and developing the best therapeutic strategies are topics of great interest. So far, DNA methyltransferases and demethylases have been successfully used in cancer therapy (91). These drugs target the entire methylation machinery in an attempt to modulate total cell transcription but in the future it might be able to target specific loci. New drugs for specific modulation have been slow to materialize, however, largely because epigenetic mechanisms and their role in gene expression are more complex than originally thought. Regardless of treatment options, DNA methylation could become useful as a biomarker for sleep deprivation. Various professionals such as pilots and truck drivers could benefit from being able to determine objectively if they have had enough sleep. Cortisol and melatonin levels are already viable for this purpose although cortisol suffers from the fact that it responds differently to acute and chronic sleep deprivation (92).
Svensk sammanfattning

Den mest välkända mekanismen för epigenetisk reglering är DNA-metylering och flera studier har undersökt kopplingen mellan metylering och övervikt. I artikel I så analyserade vi metyleringsprofilen hos 46 försökspersoner med microarray (Illumina 27K). Effekten på försökspersonernas DNA metylering studerades baserat på fetma och ålder. Artikel I visar att övervikt och fetma påverkar åldersberoende metyleringsmönster, vilket ger en möjlig molekylär länk mellan åldrande och övervikt. Detta kan visa på biomarkörer som kan användas för att förstå mekanismerna bakom åldrande, övervikt, och relaterade sjukdomar. Artikel II använde vi en annan microarray (Illumina 450K) för att undersöka effekten av gastrisk bypass, och påföljande viktnedgång, på metylering och relaterade detta till metyleringen hos friska frivilliga kontroller. Vi lyckades definiera 115 ställen i genomet där metyleringen hade ändrats efter magsäcksoperation och 41 av dessa visade sig korrelera med BMI. Bland dessa så var gener inblandade i metaboliska processer överrepresenterade (n=36, p<0.05). Dessutom var förändringarna sådana att patienterna var mer lika kontrollerna efter operationen (p<0.0001). Våra resultat från artikel II kan lägga en grund för framtida studier på effekter av magsäcksoperationer och så hjälpa till att klargöra de molekylära mekanismerna. Sömnbrist är starkt korrelerat med övervikt samtidigt som en av problemen med svår övervikt är obstruktiv sömnapné. Artikel III visar hur sömnbrist ökar portionsstorleken på ett sätt som drivs av både homeostatiska och hedoniska faktorer. Självvald portionsstorlek efter sömnbrist verkar bero på både individens hunger och vilken typ av mat som erbjuds. I artikel IV så undersökte vi hur sömndepraverings påverkar metyleringsprofilen I helblod I samma individer som I artikel IV. Artikelerna visar att sömndepravering påverkar DNA metylering på ett sätt som bade beror på och är oberoende av blodcellernas subpopulationer. Analysen visade an ING5, tillsammans med 269 andra gener var differentiellt metylerade vid sömnbrist. Detta antyder att DNA metylering är en cirkadisk process i människor.
Acknowledgements

I would like to thank functional pharmacology for giving me the chance to conduct my research. The team of Schiöth and Fredriksson represented a remarkable synergy that inspired me when I first started back in 2011. In my early days, I had the pleasure to work with Markus Sällman Almén whose apostrophe it took me too long a time to master. Christian Benedict taught me the essentials of doing clinical research from scratch. He was given a PhD student who never handled anything bigger than a yeast cell previously and hammered him into a human experimentalist. His enthusiasm never failed to amaze me and I predict a successful career for him and his research. Thanks to Jessica Mwinyi, Pleunie Hogenkamp, Galina Zheleznyakova, Lyle Wiemerslage and Samantha Brooks for the excellent collaborations I had with you. I wish to thank Sarah Voisin, Mathias Rask-Andersen, Nathalie Bringeland, and Adrian Boström for providing valuable discussion partners in higher statistics.

I owe a lot to all the people that were willing to work with me and a lot of my projects would not have been if it were not for, Britt Mari Bolinder, Elna-Marie Larsson, Karin Åberg, Arvid Morell, Eva Näslund and the others at the radiology department. Bernd Schultes and Barbara Ernst trusted me with their valuable samples, for which I am very grateful, and Janis Klovins and Davids Fridmanis have provided much needed assistance. My projects would have been equally impossible if not for students such as Victor Nils-son, Lina Lundberg, and Sanaz Zarei. Thanks to Gustav Christofferson, Mia Phillipson, and the others at MCB, and also to Josefin Jacobsson and Colin Chapman, all of which have helped me in my research.

A big thanks goes to all my coworkers at the lab, Mathias, Emelie, Wei, Linda, Mike, Lisanne, Anna, Olga, Madeleine, Sofi, Sahar, Maria, Marcus, Anders, David, Atieh, Annika, Pawel, Nataly, and all the others that I have met during my time as a project student and as a PhD student.

This would not have been possible without the two most important people in my life, Elin and Cevin. They have always been able to keep me going. Thanks to my family who have put up with an increasingly nerdy little brother. Even if you may believe otherwise, this would not have been possible without you.
References

26. . !!! INVALID CITATION !!!


A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)