Studies into epigenetic variation and its contribution to cardiovascular disease

Talens, R.P.

Studies into epigenetic variation and its contribution to cardiovascular disease

Rudolf P. Talens

Studies into epigenetic variation and its contribution to cardiovascular disease

Rudolf Talens

PhD thesis with summary in Dutch
ISBN: 978-94-6295-048-1

Front cover and internal layout by R. Velds
Printed & published by:
Proefschriftmaken.nl || Uitgeverij BOXpress, ‘s-Hertogenbosch

Studies into epigenetic variation and its contribution to cardiovascular disease

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden.
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 8 januari 2015
klokke 16.15 uur

Door

Rudolf Pieter Talens
geboren te Eindhoven
in 1976

Promotiecommissie

Promotoren:	Prof. Dr. PE Slagboom
	Prof. Dr. JW Jukema
Co-promotor:	Dr. BT Heijmans
Overige leden:	Prof. Dr. H Putter
	Prof. Dr. H Pijl
	Prof. Dr. SM van der Maarel
	Prof. Dr. DI Boomsma *Vrije Universiteit Amsterdam*
	Prof. Dr. J Mill *University of Exeter*

The research described in this thesis was supported by a grant of the Dutch Heart Foundation (DHF-2006B083) . Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

Chapter 1
Introduction to epigenetic research on common diseases

BOX1: The Hunger Winter Families Study

We explored the potential for epigenetic changes due to prenatal adversity on participants of the Hunger Winter Families Study [11]. This study investigates individuals who were exposed in utero to the Dutch Famine. The Dutch Famine was the consequence of a German imposed food and fuel embargo for the western part of the Netherlands in the winter of 1944–45, toward the end of World War II. The famine was a 6 month period of severity and extreme stress. Official daily rations contained less than 700 kcal on average, and although the winter was average meteorologically, the fuel shortages would make it tough. As the civil infrastructure was maintained during the war, individuals who were either conceived or born during the Dutch Famine, could be tracked through their birth records at the institutions in which they were born (midwife clinics in Rotterdam and Amsterdam, university hospital in Leiden. These individuals were compared with their same sex older or younger siblings. This design achieves to isolate the prenatal exposure event as best as possible in human population studies, by controlling for childhood environment and social economic status, sex, post-natal experience of the Famine and genetic variation, as siblings still share half of their genome. This study focuses on insults during the important stages of gametogenesis, organ and tissue development and the late gestational period of fetal (and neonatal) growth, characterized by remarkable increase in size.

Coronary heart disease (CHD) is an ageing related common disease that was responsible for 83,941 hospitalizations and 9,720 deaths in the Netherlands in 2012 [1]. It is caused by a narrowing of the coronary blood vessels with a myocardial infarction (MI) being its most common manifestation. CHD is a disease of a complex multifactorial nature. An individual's risk of CHD is determined by the lifelong interaction between genetic background and environmental exposures [2,3]. Genetic and environmental factors contribute equally to CHD etiology [4]. Investigations into genetic sequence variation have been a main focus of the past decade and have revealed many genes behind the heritable components of CHD and its risk factors (e.g. type 2 diabetes (T2D) complications, metabolic syndrome, or hypertension) [5]. Classical environmental risk factors of CHD (e.g. smoking, diet, low physical activity) are well-established as contributors to its clinical pre-stages such as hypertension, lipid profile, and obesity [6]. The risk of CHD is strongly correlated with age [1,7], which may reflect an accumulation of environmental exposures and ageing-related (stochastic) processes that both result in a randomness in onset and etiology of CHD. Geographical and epidemiological

studies showed that poor conditions at the beginning of life also contribute to risk and severity of CHD in later life [8,9]. A survey across studies on famine cohorts, amongst which the Dutch Famine of 1944-1945 [Box 1], reported consistent associations between prenatal famine exposure and a higher BMI, more adiposity, a less favorable glucose metabolism and a more adverse lipid profile [10].

While the identification of genetic risk factors of CHD has contributed insights in molecular pathways involved in its etiology, the molecular mechanisms that mediate the effect of environmental risk factors are less well understood. Intervention studies on animal models present evidence that changes in DNA methylation, a major epigenetic mark, may be mechanistically involved in the association of prenatal adversity and ageing-related disease phenotypes [12,13]. In humans changes in DNA methylation at the *IGF2* locus were associated with exposure to the Dutch Famine during conception [14]. Changes in DNA methylation have been repeatedly found associated with lifestyle and environmental exposures [15]. Ageing has been repeatedly associated with stochastic changes in DNA methylation and other epigenetic marks such as histone modifications [16]. Thus, research on ageing related common diseases such as CHD has partly refocused its attention towards the involvement of epigenetic regulation, as a mediator of the interactions between genome and (prenatal) environment [17,18]. This general introduction will explain the biochemical basis of epigenetic mechanisms, explore current knowledge on its involvement in disease development, and describe the facilities and resources available to epigenetic research, and finally present an outline of this thesis.

Epigenetic mechanisms

The genome of multi-cellular organisms appears in a condensed state within the nucleus, forming a nine to thirty nanometer thick fiber of a DNA – protein complex called chromatin [19,20]. These chromatin fibers are formed by the coiling and packaging of its nucleosomes, which consist of 147 base pair long stretches of DNA wrapped twice around a protein scaffold of eight different histone subunits. Classically the condensation state of chromatin is grouped in three structural categories, open (euchromatin), intermittent (bivalent chromatin), and closed (heterochromatin), although

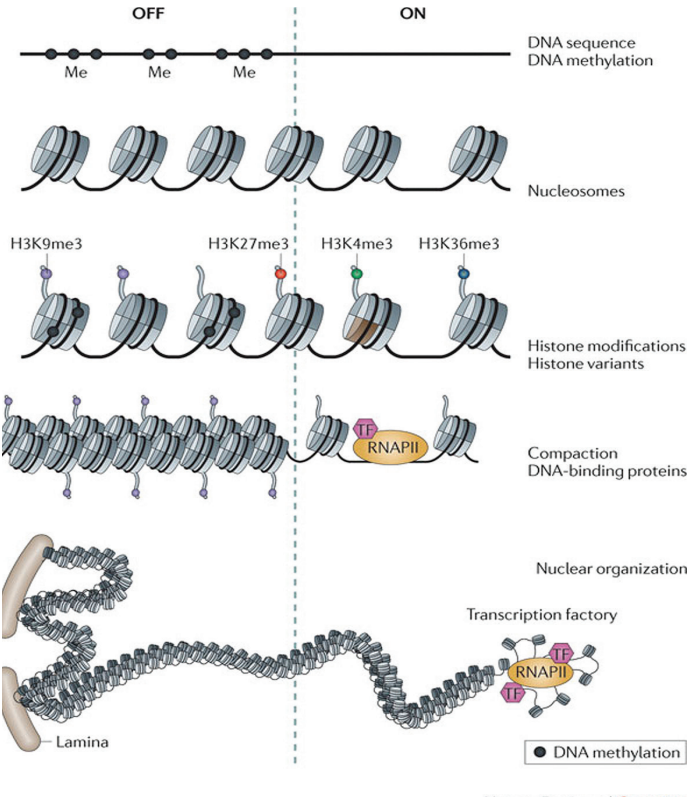


Figure 1.1 The different layers of epigenetic information, exemplified by the two extremes of the spectrum: transcriptionally inactive heterochromatin (left) and transcriptionally active euchromatin (right) with a selection of their associated epigenetic marks. The primary layer of chromatin structure constitutes the methylation of genomic DNA (Me) at cytosine bases in specific contexts, and the packaging of DNA into nucleosomes, which vary in histone composition (differently colored pie slices) and modifications of the histone tails (indicated by colored dots). DNA in chromatin may remain accessible to DNA-binding proteins such as transcription factors (TFs) and RNA polymerase II (RNAPII) or may be further compacted. Chromatin can also organize into higher-order structures such as nuclear lamina-associated domains and transcription factories [21]. *Image reprinted with permission from Macmillan Publishers Ltd: Nature Reviews Genetics Volume 12 (1); Zhou, Goren & Bernstein, pages 7-18; copyright 2011*

more distinct chromatin states were recently defined [21]. The amount of condensation at each genomic area is regulated by epigenetic mechanisms, the molecular basis of which is formed by several correlated layers of biochemical information on and around the nucleosomes (Figure 1.1).

These epigenetic layers include the methylation of DNA at cytosine residues, the addition of small chemical groups (e.g. phosphate, acetyl, or methyl groups) to the various histone subunits, nucleosomal packaging; non coding RNAs; and nuclear localization [22–25].

The capacity of the transcription machinery to access a gene after receiving the correct cues decreases with increasing condensation of its chromatin. A further assortment of epigenetic modifications, mainly on histone tails, results in the attraction or repulsion of transcription factors (i.e. enhancers and repressors) that enables a more or less continuous spectrum of fine tuning for epigenetically regulated gene expression [21,26]. Epigenetic silencing of repetitive DNA is generally thought to protect our genomes against the potentially detrimental effects of random insertion of transposons, mobile elements of parasitic origin [27,28]. Further, through variation of the epigenetic state at a genomic locus, epigenetic mechanisms determine the gene expression potential of a cell without changes to its DNA sequence [29]. Extensive epigenetic remodeling during development and differentiation generates the multitude of our body's cell types, each with its unique cellular phenotype, function, and response repertoire, that nonetheless all contain the same genotype as the zygote they ultimately derived from [30].

A textbook example illustrating the importance of epigenetically regulated gene expression is the mechanism of genomic imprinting [31,32]. Imprinted genes are epigenetically marked during gametogenesis so that they are expressed only from the paternally or maternally derived chromosome. Only about 0.3 % of our genes are imprinted, most of such genes reside in clusters, such as the *GNAS* locus which contains silenced genes of both maternal and paternal origin and at least one non coding RNA that regulates imprinting of the adjacent genes [33]. Most imprinted genes code for key players in the signaling pathways that coordinate growth and development with metabolism [34,35]. Their effect covers as diverse processes as the development of extra-embryonic tissues, fetal nutrient acquisition, neonatal suckling, and the control of appetite during life [33,34]. Incorrect establishment of imprinting marks leads to fetal lethality, growth impairments, insatiable appetite, and mental retardation [34], exemplified by imprinting disorders

such as the antagonistic Prader-Willi_Angelman syndrome [36], and Beckwith-Wiederman_Silver-Russel syndrome [37].

In contrast with the genetic sequence, epigenetic marks are not permanently fixed in a differentiated cell, they are, however, usually maintained over longer periods of time, and are faithfully transmitted to its daughter cells. During cell division the DNA separates from its protein scaffold to split the duplex for replication. Although several mechanisms have been proposed [38], mitotic transmission of the various histone modifications is less well understood than that of DNA methylation. The methylation marks are directly transmitted, as the methyl groups are covalently bound to the old strands of DNA, each of which is passed to one daughter cell. DNA methylation is the most studied layer of epigenetic information. A high degree of DNA methylation is generally associated with heterochromatin (Figure 1.1) and was long considered an epigenetic mark specific for the repression of transcription [39], although recent results from genome-wide techniques show that its function depends on context of its location within the transcriptional unit [28].

In humans, most DNA methylation occurs at the cytosine residue of CpG sites (a cytosine directly followed by a guanine residue) [24], but this is not the only form of DNA methylation. In embryonic stem cells and induced pluripotent stem cells substantial DNA methylation has been observed outside a CpG context, which is removed during differentiation [40,41]. Two recent studies have discovered cytosine residues that had a hydroxy-methyl group (hmC) covalently bound instead of a methyl group (mC) in both a-mitotic cell types of the brain [42] and in embryonic stem cells [43]. In stem cells (i.e. embryonic, pluripotent, and induced) high levels of hmC are often observed at epigenetically poised loci involved in pluripotency and maintenance of differentiation potential. The finding that methyl CpG binding proteins have a reduced affinity for hmC compared with mC [44], suggests that during differentiation the conversion of mC to hmC may be a molecular mechanism for rapid epigenetic decisions [43,45]. Its proposed role as biochemical intermediate in the process of active and passive DNA de-methylation [43,45] further supports the importance of hmC in cell differentiation, but also holds that it is likely not retained during multiple cell divisions. In line with this, hmC was found to occur only in low to marginal levels in tissues

relevant for CHD development [46]. Thus, most studies on the epigenetic contribution to CHD and other common ageing related diseases focus on the methylation of DNA at CpG sites in differentiated tissues.

The epigenetic contribution to diseases with an environmental component

The regulatory role of epigenetic mechanisms in cellular gene expression has forwarded research on the epigenetic involvement in common diseases, to both complement and understand their genetic component [47], with promising results. Two recent longitudinal studies reported an association of developing obesity [48] or type 2 diabetes [49] with different DNA methylation levels in leukocytes at established genetic candidate loci. Both studies reported that the epigenetic differences were found to precede disease manifestation and had remained stable throughout disease development during the study, showing that the epigenetic variation was persistent [48,49]. The epigenome has a dynamic capacity to undergo changes in each individual cell. This potential changeability of epigenetic marks is another topic of interest in research on common diseases [50]. The most studied example of epigenetic changes involved in disease are those affecting tumor cells [51]. Tumor cells often display high DNA methylation at promoters of genes controlling cell cycle arrest, while the genes promoting cell division are unmethylated, and both changes have been associated with the uncontrolled growth that is typical of malignant tumors [52–54]. Epigenetic differences have also been observed in ventricular tissue affected by MI compared with unaffected tissue from the same individual [55,56]. Studies on monozygotic twin pairs that are discordant for Beckwith-Wiederman and Silver-Russel syndromes suggest that epigenetic changes can precede the manifestation of metabolic diseases [57], and a recent genome wide study reported associations of DNA methylation with calendar age and ageing related phenotypes [58]. These and other results showcase that epigenetic variation may be an important factor in the process of ageing and risk for its related common diseases [18,59].

Individuals may harbor persistent epigenetic differences due to, first of all, the genetic sequence, which may prevent or force the establishment of epigenetic marks [60,61].

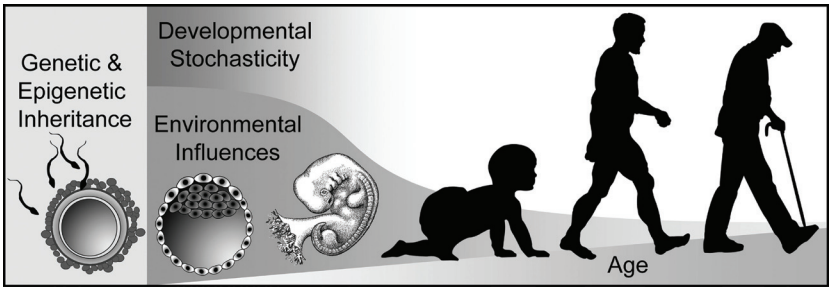


Figure 1.2 Contrary to genetic sequence variation, which is fixed for life after fertilization, epigenetic variation has a more changeable nature that involves different factors during phases of life. Two periods of prenatal life are proposed to be particularly vulnerable to stochastic and environmentally driven changes. First, the conditions in which both parents live prior to fertilization can disturb the extensive epigenetic remodeling during gametogenesis. Genetic sequence variation may form another source of epigenetic variation (e.g. polymorphisms affecting CpG sites) in the zygote. Second, the epigenetic remodeling during tissue development can be shaped by both stochastic factors (e.g. epigenetic differences in MZ twins), and by the intra-uteral environment, shaped by both internal physiological factors (e.g. placental morphology) and external maternal environmental factors (e.g. smoking, nutrition). Cell type differentiation also involves epigenetic remodeling, especially during early post-natal growth, resulting in a potentially still increased susceptibility to childhood environmental exposure like nutrition. During adult life environmental factors (lifestyle, living conditions) are thought to be of equal importance to stochastic changes (due to imperfect epigenetic maintenance), while in old age the efficiency of epigenetic maintenance is thought to decrease [119]. *Immage republished with permission of Annual Reviews from Annual Review of Nutrition, volume 27, pages 363 – 388: Epigenetic Epidemiology of the Developmental Origins Hypothesis; Waterland & Michels, copyright 2007; permission conveyed through Copyright Clearance Center, Inc.*

Next, epigenetic differences between newborn monozygotic twins [62] and isogenic littermates [63] indicated that the extensive developmental program of epigenetic remodeling [30] may induce stochastic epigenetic variation, as has been proposed [64]. Further, a wealth of intervention studies on animal models showed that prenatal adverse conditions can induce epigenetic changes [65,66]. In humans, prenatal famine [14], and smoking during pregnancy [67] were also found associated with persistent epigenetic differences. Early development may provide a period of specific vulnerability to environmental influences (Figure 1.2) after which the epigenetic differences may persist through faithful transmission during mitosis [68,69].

In differentiated cells, epigenetic information can change

over time. Imperfect maintenance of epigenetic marks has long been proposed to result in stochastic changes without any environmental influence [50,70]. A recent study reported solid evidence for this by following two initially identical cell cultures growing separately under the same conditions for 300 generations [51]. Epigenetic information may further change under the influence of environmental exposures related to lifestyle and living conditions [15] (Figure 1.2). For instance, prolonged or repeated exposure in adulthood to tobacco smoke [71,72], traffic exhaust fumes [73] or low doses of arsenic [74] were found associated with differences in DNA methylation. Intrinsic cues such as chronic inflammation [75,76] or stress [77] were also found to induce postnatal epigenetic changes.

Epigenetic changes due to environmental exposures could result from direct chemical damage, or from the incomplete (e.g. lack of methyl donors) or compromised (e.g. toxic interference) establishment of epigenetic marks. It is even conceivable that environmentally driven epigenetic changes may to some extent constitute an adaptive response [65], as repeated transcription of a locus has been suggested to change its epigenetic state [21,26]. Changing epigenetic marks could thus be an indicator of an active, or activated biological process, although their downstream biological effects remain to be revealed [15]. In conclusion, at birth epigenetic variation may be persistent, semi-heritable, and to a degree soma-wide [69]. Then, as uncorrected stochastic aberrations and environmentally driven epigenetic changes accumulate throughout life, this persistent level gradually changes in each cell individually [15]. It has been hypothesized that a build-up of uncorrected dysfunctional epigenetic perturbations may eventually result in a state called epigenetic dysregulation, which is implicated in ageing and its related diseases [58,59].

Technology, data resources, and biobanks in epigenetic research

The huge advance in DNA sequencing technologies of recent years (reviewed by Harrison & Parle-McDormott (2011) [78]) has enabled measurement of DNA methylation at individual CpG sites on both genome wide and high throughput platforms [Box 2]. Genome wide platforms are best suited for constructing epigenome maps and for

BOX2: Measuring DNA methylation by sequencing bisulphite treated DNA

Most studies assess DNA methylation by sequencing DNA that is treated with bisulphite (BS), although more techniques exist (reviewed by Harrison & Parle-McDormott (2011) [78]). BS treatment transforms unmethylated cytosines into uracils, while retaining the methylated cytosines, creating an artificial sequence variant that genetic sequencing tools and techniques can measure. However, BS based methods cannot distinguish between methyl-cytosine and hydroxy-methyl-cytosine [81], an issue in stem cells and brain, but not in other somatic cells [46]. Determining the relative proportions of the artificial BS sequence variants on a representative number of DNA molecules, which requires a slight alteration of the protocols from genetics, forms a measure for the amount of methylated and unmethylated alleles in the tissue sample. The DNA sequencing technologies developed in recent years have been modified to enable sequencing of BS treated DNA, and thus measurement of DNA methylation at individual CpG sites on whole genome, genome-wide and locus-specific medium throughput platforms (reviewed by Gupta et al. (2010) [82] and by Harrison & Parle-McDormott (2011) [78]). An example of a platform for whole genome BS sequencing is the powerful Illumina HiSeq platform (Illumina, San Diego, USA), which uses next generation sequencing technology to assess the methylation of over 90 % of the genomic total of 28 million CpG sites [83]. An example of a genome-wide platform is the Infinium BeadArray platform (Illumina), which has increased the amount of CpG sites it can interrogate from 1.500 in 2006 [84] to over 450.000 in 2011 [85]. Examples of medium throughput platforms are the sequencing-by-synthesis based Pyrosequencing platform [86] and the mass spectrometry based Sequenom MASSarray platform [87].

uncovering epigenetic candidate loci. However, these platforms still interrogate only a fraction of all CpG sites in the human genome and are biased towards CpG islands and promoter regions [79]. Medium-throughput platforms, like the Sequenom EPITYER platform, enable accurate quantitative measurement of methylation of a limited set of (nearby) CpG sites at any genomic (candidate) locus on large populations [80].

Following the technological aspect of measuring data, epigenetic research will need resources to facilitate interpretation of epigenetic associations. Parallel to genetics, these resources must include epigenome maps charting consensus epigenetic marks (DNA methylation and histone modifications), preferably for multiple cell types [88,89] and epivariome maps describing inter-individual variation in these marks [90–92]. Further, notwithstanding a few notable exceptions [93], the regulatory effect of DNA methylation in a cell population is commonly thought to rely on the status of multiple CpG sites within a genomic region [94], requiring data describing such potentially exploitable patterns within the normal variation (cf. linkage disequilibrium) [60,95]. Epigenetic databases, like those from the International

Human Epigenome Consortium [96] (http://www.ihec-epigenomes.org/) and the National Institute of Health (NIH) Epigenomics Roadmap [97] (http://www.roadmapepigenomics.org/), are under construction, but will require much effort to be as effective as their genetic counterparts.

Epigenetic research will also require access to diverse biomaterial of human population or patient based studies that investigate ageing and its common diseases. Many of these studies have collected blood samples at baseline, subsequently storing the extracted DNA in biobanks for future investigation. Epigenetic studies on these biobanks are technically possible since DNA methylation is not affected by standard methods for DNA extraction and storage. However, these biobanks were originally set up for genetic research and often contain only DNA from blood obtained during a single round of sampling. Their use in epigenetic research may thus be limited, since their construction did not consider resampling to allow studies on a dynamic epigenome [94]. First, epigenetic marks are susceptible to environmental and physiological influences. This may impede cross-sectional epigenetic studies in distinguishing cause from consequence of disease [98,99], while prospective epigenetic research will require either longitudinal sampling [49] or another form of evidence that the epigenetic variation remained stable during the study. Second, the frequently described tissue differences in (genome wide) epigenetic patterns [100–102] may limit the relevance of epigenetic variation to the tissue that was assessed, which may confine most epigenetic studies into disease causes to the search for potential epigenetic markers of disease risk, and limit implicating an epigenetic mechanism in disease etiology to those diseases for which the relevant tissues are available. Third, and more pressing is the fact that inter-individual variation in the composition of the heterogeneous leukocyte population may confound the epigenetic variation in blood, since each different cell type may carry its own specific epigenetic marks [103]. Although not relevant for candidate gene studies, methods that address this issue through mathematical correction are being developed for genome wide investigations [104].

Unfortunately there will never be biobanks that contain repeated longitudinal DNA samples from disease related internal tissues (e.g. heart, brain), thereby dismissing the direct analysis of epigenetic marks on other than biosamples

Table 1.1: Candidate loci investigated in the various studies of this thesis

Locus	Chromosome	Gene function	Studied in chapter				
			2	3	4	5	6
IL10	01q32.1	Anti-inflammation	x	x		x	
NR3C1	05q31.3	Stress response	x	x			
TNF	06p21.33	Pro-inflammatory	x	x			
IGF2R	06q25.3	Growth/Apoptosis	x	x			
GRB10	07p12.2	IIS inhibitor	x	x			
LEP	07q32.1	Metabolism	x	x	x	x	
CRH	08q13.1	Stress response	x	x	x		
ABCA1	09q31.1	Cholesterol transport	x	x	x	x	
IGF2[a]	11p15.5	Early growth	x		x	x	
INSIGF	11p15.5	(Embryonic) Growth	x	x	x	x	
KCNQ1OT1	11p15.5	Imprinting control region	x	x	x		
MEG3	14q32.2	Growth suppressor	x	x			
FTO	16q12.2	Development	x	x			
APOC1	19q13.32	Metabolism	x	x			
GNASAS	20q13.32	Growth/Lypolytic signal	x	x	x	x	
GNAS A/B	20q13.32	Growth/Lypolytic signal	x	x			

a: The susceptibility of DNA methylation to prenatal conditions was observed previously at the *IGF2* locus [14,120].

obtained during post mortem examinations [105] or repeated sampling of peripheral tissues. However, loci at which epigenetic marks are set early in development, for instance imprinted loci, are commonly thought to harbour a similar epigenetic information throughout the soma. Thus, at such loci epigenetic variation in peripheral tissues is proposed to also inform on internal tissues [106]. Although not often collected, other peripheral tissues than blood can be obtained with minimal (e.g. buccal cells, or epidermis) [107,108] or moderate invasiveness (e.g. fat, or muscle biopsies) [109,110], which may be useful in elucidating similarities of epigenetic variation across tissues. In recent years a few projects have been launched to establish biobanks designed for epigenetic research that for now mostly contain cross-sectional samples from multiple peripheral tissues [62,111]. However promising the emergence of such projects, it will take time for many more of them to emerge. Thus, if the clinical biobanks with all their limitations can still provide meaningful answers to some epigenetic research questions,

they will present an unmatched source of patient based DNA samples, due to their combined size, disease repertoire, and detail of participant information gathered during the (long) follow up period.

Outline of thesis

Epigenetic marks can be persistently or dynamically altered from gametogenesis until death by a combination of influences from exposure to the external environment, from internal physiological processes, and from uncorrected stochastic perturbations. These sources of epigenetic variation cannot be readily isolated for investigation, particularly not in humans, nor will it be possible to capture a lifetime's accumulation of causes and consequences of epigenetic variation in one study. However, careful formulation of specific research question and subsequent selection of appropriate study designs, may still promise meaningful epigenetic research on ageing and its related diseases [112]. In this thesis we will therefore apply a sequence of studies to investigate the relation between epigenetic mechanisms, ageing, and CHD.

Chapter 2 investigates properties of epigenetic variation at 16 candidate loci, that were selected from the domain of cardiovascular and metabolic diseases (Table 1.1), pertaining to the suitability of DNA samples from existing biobanks for epigenetic studies on these loci. First we describe the inter-individual variation and patterns of DNA methylation at 104 individual CpG sites of the loci on recent blood samples from unrelated, healthy participants (14 – 72 years old) of the Netherlands Twin Register (NTR) [113,114]. We also assess the influence of leukocyte population heterogeneity on this variation and present a method to account for its potential effect on epigenetic variation. Then, we determine the temporal stability of methylation variation on longitudinal DNA samples from blood (mesoderm) and buccal cells (ectoderm) over a time span of up to two decades and we survey its correlation between the recent samples of both tissues.

Chapter 3 investigates the effect of prenatal exposure to the Dutch Famine of 1945 on epigenetic marks in individuals from the Hunger Winter Families Study [Box 1] [11]. We compare DNA methylation at 15 of the candidate loci (Table 1.1) between individuals (57 – 59 years old) who were prenatally exposed to the Dutch Famine and their same sex younger or older siblings

(43 – 70 years old), who were not exposed in utero. By measuring middle aged individuals we dismiss transient epigenetic changes that would reverse with improving conditions, thus focusing on lasting effects. The uncovered associations between prenatal famine and persistent DNA methylation are further explored for the role of the timing of the exposure (around conception vs. last trimester) and sex of the exposed.

Chapter 4 investigates the extent to which genomic DNA methylation, approximated with the LINE1 assay [115], and DNA methylation at 7 of the candidate loci (Table 1.1) changes stochastically or due to environmental influences during the adult life span. We assess inter-individual and within-pair variation in DNA methylation with increasing calendar age in a combined cross-sectional and longitudinal study on a large population of Dutch and Danish monozygotic twin pairs (18 – 88 years old) [116,117]. We also evaluate the influence of familial versus individual factors on age-related epigenetic variability. The twins pairs are solely selected on their age in order to randomize potential differences in adult living environments, thus focusing the study on the amount of epigenetic changeability across the full adult lifespan due to the combined effects of stochastic and environmentally driven epigenetic changes.

Chapter 5 investigates whether DNA methylation at the 6 loci at which it marks prenatal environmental conditions (Table 1.1**)** is associated with the risk for developing a myocardial infarction (MI) in a population of individuals (70 – 82 years old) from the PROSPER study [118]. The individuals selected from this study had not experienced a previous CHD event in order to exclude potential epigenetic differences as a consequence of the disease.

Chapter 6 explores the extent to which the variation of DNA methylation in blood may inform on the methylation variation in internal tissues using a genome wide survey of post mortem samples of blood, subcutaneous fat (SC fat), muscle, visceral fat (VS fat), liver, spleen, and pancreas obtained from six individuals (58 – 79 years old) during autopsy. DNA methylation is assessed at 378,239 CpG sites distributed throughout the autosomal chromosomes. Besides comparing the tissues on genome wide patterns, we assess differences in the methylation level and the amount of variation at individual CpG sites and we inspect correlation coefficients between the tissues. Pooled blood from healthy middle aged regular blood donors is included in the study, to ascertain that no overhauling of DNA methylation patterns occurs within the first 24 hours after death.

Chapter 2

Variation, patterns and temporal stability of DNA methylation: considerations for epigenetic epidemiology

Rudolf P. Talens[1], Dorret I. Boomsma[2], Elmar W. Tobi[1], Dennis Kremer[1], J. Wouter Jukema[3], Gonneke Willemsen[2], Hein Putter[4], P. Eline Slagboom[1], Bastiaan T. Heijmans[1,*].

1. Leiden University Medical Center, Department of Molecular Epidemiology, Leiden, The Netherlands
2. VU University Amsterdam, Department of Biological Psychology, Amsterdam, The Netherlands
3. Leiden University Medical Center, Department of Cardiology, Leiden, The Netherlands
4. Leiden University Medical Center, Department of Medical Statistics, Leiden, The Netherlands

Published in FASEB Journal, Volume 24, Issue 9, pages 3135 – 3144; 2010

Abstract

The prospect of finding epigenetic risk factors for complex diseases would be greatly enhanced if DNA from existing biobanks, which is generally extracted from whole blood, could be used to perform epigenetic association studies. We characterized features of DNA methylation at 16 candidate loci, 8 of which were imprinted, in DNA samples from the Netherlands Twin Register biobank. Except for unmethylated or fully methylated sites, CpG methylation varied considerably in a sample of 30 unrelated individuals. This variation remained after accounting for the cellular heterogeneity of blood. Methylation of CpG-sites was correlated within loci and across chromosomes for 4 imprinted loci. In 34 additional individuals, we investigated the DNA methylation of 8 representative loci in 2 longitudinal blood and 2 longitudinal buccal cell samples (follow-up 11-20 and 2-8 years, respectively). 5 of 8 loci were stable over time ($\rho>0.75$) in both tissues indicating that prospective epigenetic studies may be possible. For 4 loci, the DNA methylation in blood (mesoderm) correlated with that in the buccal cells (ectoderm) ($\rho>0.75$). Our data suggest that epigenetic studies on complex diseases may be feasible for a proportion of genomic loci provided that they are carefully designed.

Key words

Epigenome; Biobank; Human studies; Complex disease

Introduction

Epigenetics refers to heritable differences in gene expression potential that are not caused by variation in the DNA sequence [29,119]. Its molecular basis is the chemical modification of either the DNA itself (cytosine methylation in CpG dinculeotides) or the histones that package the chromatin (e.g. methylation, acetylation, phosphorylation.) [22,24,121]. It has frequently been proposed that changes in these epigenetic marks significantly contribute to the risk of complex diseases including cancer, cardiovascular and metabolic disease [17,98,119,122–124]. However, with the exception of studies on cancer, empirical data from epidemiological studies supporting these hypotheses are largely absent mainly due to technical and methodological limitations.

Many of the technical limitations have been resolved, in particular with respect to the high throughput measurement of DNA methylation [125,126]. DNA methylation is correlated with other layers of epigenetic marks, particularly histone modifications [23]. DNA methylation may be the most suitable epigenetic mark for large-scale epidemiological studies, since methyl groups are covalently bound to CpG dinucleotides and are not lost during routine DNA extraction, unlike histone modifications. This opens the possibility of exploiting existing DNA biobanks for research purposes, to discover epigenetic risk factors for complex disease.

Epigenetic studies will require the development of data resources analogous to those that facilitated genetic association studies. The resources should include epigenome maps charting DNA methylation marks [41], the description of inter-individual variation in DNA methylation (cf. single nucleotide polymorphisms and copy number variants) [127] and data on the patterns within this variation (cf. linkage disequilibrium)[60]. To guide the development of such epigenome-wide resources, candidate loci may be studied. In this respect differentially methylated regions influencing imprinting [128], transposon-derived sequences [128], CpG island shores [129] and recognition sequences for methylation-dependent transcription factors [93] are of particular interest.

In addition, several issues potentially limiting the use of existing biobanks for epigenetic epidemiology need to

be addressed. Firstly, DNA in biobanks is mostly extracted from whole blood, which, like any tissue, consists of different cell types that may carry different epigenetic marks and whose relative numbers may vary between individuals [103]. Secondly, the stability of DNA methylation over time should be known before the association of DNA methylation with future disease risk can be assessed. Global (or average) DNA methylation has been reported to change over time [130,131], but DNA methylation of specific loci may be more stable [60,132]. Lastly, it will be crucial to address to what extent DNA methylation measured in blood marks in less accessible tissues that are directly involved in disease. Despite scattered reports that this may be the case [100,133,134], the issue remains largely unresolved.

We assessed whether genomic DNA stored in existing biobanks would be suitable for epigenetic epidemiological studies. To this end we addressed the inter-individual variation in DNA methylation of 16 candidate loci for cardiovascular and metabolic disease, the influence of blood cell heterogeneity on this variation, the stability of DNA methylation over time and its correlation between whole blood (mesoderm) and buccal cells (ectoderm) in individuals from the Netherlands Twin Register (NTR) [113,114].

Methods and Materials

Study populations

The individuals investigated in this study were selected from the Netherlands Twin Register (NTR) biobank [113,114], which includes DNA samples from Dutch twins and their family members (parents, siblings, offspring and spouses). Firstly, unrelated individuals (Table S2.3) were selected (n=30) to study inter-individual variation in DNA methylation, the influence of cell heterogeneity and patterns of DNA methylation. These 30 individuals were selected from the ongoing NTR biobank project for which 9560 individuals were included. Random selection would result in a sample of individuals with characteristics very close to the average in the complete cohort. For this we applied the D-optimality criterion to the Fisher information matrix, which enabled us to select 30 individuals, representive of the whole range of phenotypic variation in age and metabolic parameters

present in the complete cohort. The age of the individuals selected ranged from 21 to 73 years; metabolic parameters of interest included: waist circumference, fasting blood glucose level, serum LDL and HDL cholesterol. Plasma and serum measurements and cell counts of whole blood were obtained using the standardized methods previously described (missing cell count information for 2/30 individuals). Furthermore, we selected the proportion of males to females and of those who had never been smokers, to former and to current smokers so that it was equal to the proportion in the complete cohort.

Secondly, 34 individuals were selected for assessing the correlation of DNA methylation across time and tissue, and for validating the findings on the group of 30 individuals described above. This group consisted of participants in the NTR biobank project, who also took part in previous NTR projects. This allowed for recent DNA samples from whole blood (with information of cell counts) and buccal cells, as well as previous DNA samples from whole blood drawn 11 to 20 years earlier, and from buccal swabs taken 2 to 8 years earlier (Table S2.4). The age at first sampling ranged from 14 to 62 years. Among the 34 individuals, 17 were male, 26 were unrelated individuals and 8 were monozygotic twin pairs. DNA from all samples was extracted from whole blood and buccal swabs using standard methods.

Thirdly, the results on within-individual correlation between CpG-units were validated using 60 controls (28 males, mean age 57 years) from the Dutch Hunger Winter Families Study [11]. DNA methylation was measured at the same loci using the same methods as the current study [14,135].

DNA methylation

Loci were selected on the basis of their potential involvement in cardiovascular and metabolic disease through the role of the adjacent candidate gene in growth, lipid metabolism, energy metabolism, inflammation, or stress response. Assay design focused on the regions of these loci that contained features with a potential for epigenetic regulation as observed in human, animal or cell culture experiments [93,133,136–148]. The loci selected included promoter elements, CpG-islands, transposon-derived sequences, methyl-sensitive transcription factor binding sites

(mTFBS), imprinted differentially methylated regions (DMR) and regions reported to regulate transcription through DNA methylation. Methylation assays were designed using the methprimer tool [149] on sections of sequence downloaded from the UCSC genome browser [150]. 58 Assays were tested for the reliability of the methylation measurement. 40 Assays gave a reliable measurement, and based on the priority given to the associated candidate gene and epigenetic properties (Table 2.1), 16 of these were selected to cover the whole range of possible average methylation levels (0%-100%).

One microgram of genomic DNA was bisulfite-converted using the EZ 96-DNA methylation (Zymo Research). DNA of the 30 individuals, in whom variation in DNA methylation was investigated, was converted on a single 96-well plate. DNA methylation of all 16 loci was measured using the same bisulfite-converted sample. The 4 samples (blood, buccal, recent and old) from individuals selected for testing the correlation over time and across tissues were bisulfite-treated on the same 96-well plate. For this sub-study, two 96-well plates were used to process the 136 samples, each plate with an equal number of individuals. Methylation of the 8 loci was measured using a single bisulfite-converted DNA sample. Primers used to amplify the region of each assay are given in table S2.5A. DNA methylation was measured using a mass spectrometry-based method (Epityper, Sequenom)[87] whose quantitative accuracy (R^2 duplicate measurements $\geq$ 0.98) and concordance with clonal PCR bisulfite sequencing was reported previously [151,152]. All measurements were done in triplicate. Quality control consisted of several steps. CpG site containing fragments that had equal or overlapping mass, making them irresolvable by mass spectrometry, and CpG-sites containing fragments whose measurement was confounded by single nucleotide polymorphisms [60], according to dbSNP build 128 were discarded (Table S2.6). Next, at least two of the three replicate measurements had to be successful and the SD of the replicate measurements had to be 0.10 or less. Only CpG sites with a success rate greater than 75 % for the latter two criteria were considered fit for further analysis and the average was calculated for the replicate measurements. With these criteria applied, DNA methylation of 164 CpG-sites, distributed over 104 CpG site containing fragments (CpG units [87]), could be measured in the first sample of 30 individuals (Table S2.5B), 62 CpG units

contained 1 CpG site, 28 CpG units contained 2 CpG sites, 10 CpG units contained 3 CpG sites and 4 CpG units contained 4 CpG sites. The methylation of CpG sites occurring on a single fragment (CpG unit) cannot be resolved individually. Average CpG methylation for these CpG units was calculated using the RSeqMEth module [151]. The average success rate for the 104 CpG units assessed was 97%. In the second group of individuals, 41 CpG units containing 55 CpG sites could be measured applying the same criteria only to the recent blood samples and 38 CpG units containing 52 CpG sites to all 4 samples (Table S2.5C).

To exclude the influence of DNA sequence variation not present in dbSNP on higher correlations observed between DNA methylation measured in a recent and in an old sample and blood and buccal swab, the evaluateSNPs() function of the R-module MassArray was used [153]. The Epityper method for DNA methylation measurements is based on a protocol to resequence genomic DNA using mass spectrometry (MassCleave, Sequenom [154]). By comparing the mass spectrum observed with the one expected according to a reference sequence, data points can be identified that are suspected of being confounded by sequence variation. This interference can be directly because the sequence variant affects a fragment containing a CpG site or indirectly because a sequence variant changes the mass of a non-CpG fragment so that it overlaps with a CpG-containing fragment. It is noteworthy, that sequencing genomic DNA would deal with direct interference only. Data-points suspected to be affected by unknown sequence variation were excluded and correlations recalculated to examine their influence. CpG measurements were removed for *APOC1*, CpG 1 (4 individuals excluded for all DNA samples (recent, old, blood, buccal)), CpG 10 (8) and CpG 11 (11); *LEP*, CpG 8 (1); *IGF2*, CpG 6&7 (21); *IGF2R*, CpG 4&5 (1) and CpG 11-13 (4); for *CRH* no measurements were excluded.

Statistical analyses

Male and female data were analysed separately in every test. However, in view of the study size the outcomes of these analyses should be considered as purely explorative. All P-values are two-sided and statistical analyses were performed using SPSS 16.0.

Variation in DNA methylation.

CpG sites with average methylation levels close to 0% or 100% by definition have a truncated variance. In order to circumvent this problem, the following variance-stabilizing transformation was applied [155]. Transformed value = Arc tan ((Methylation / (1-Methylation))^2). Using the transformed values, the equality of the variance of CpG sites was tested with Levene's test.

Accounting for cellular heterogeneity.

To test whether variation in DNA methylation was confounded by cellular heterogeneity, nested linear mixed models [156] were applied to the transformed data. The basic model was created as a baseline, to be subtracted from the nested models. It included the CpG site as a fixed effect. The three nested models each had one percentage of a major white blood cell subclass, namely neutrophils, lymphocytes or monocytes, added to the basic model as an extra fixed effect to test whether variation in this percentage could explain part of the variation in DNA methylation. The actual amount of variation in DNA methylation that could be explained by the percentage of the white blood cell type was calculated as 100% minus the percentage of the residual variance of the nested model with respect to the residual variance of the basic model. The linear mixed model accounts for correlated methylation within individuals and deals with methylation data missing at random without imputation [14]. It may be seen as an extension of the paired T-test: the model will reduce to a paired T-test with identical results if between-group methylation differences are assessed for a single CpG site and if data are complete and all other factors are omitted.

Correlation and patterns of CpG methylation.

Bivariate Pearson correlation coefficients between CpG sites were calculated after transforming the transformed methylation levels for neutrophil percentage. Non-significant correlations were treated as if there were no correlation (value of correlation set to 0). Patterns in the correlation matrix were visualized with a heat map after unsupervised complete linkage clustering, which utilizes the Euclidean distance (the difference between two points in the matrix squared).

Stability over time.

The difference between DNA methylation at two time points was calculated per individual for each CpG unit as methylation of the old sample minus methylation of the new sample. Missing values were excluded pair-wise. Spearman's rank correlation coefficient was used to calculate the correlation between the two time points.

Correlation between tissues

The difference between DNA methylation in blood and buccal cells was calculated per individual for each CpG unit as methylation of the recent blood sample minus methylation of the recent buccal cell sample. Missing values were excluded pair-wise. Spearman's rank correlation coefficient was used to calculate the correlation between the two tissues.

Results

Inter-individual variation in DNA methylation

We quantitatively measured the DNA methylation of 16 candidate loci using DNA samples from the NTR biobank, to estimate its variation in the general population. The DNA was extracted from whole blood. The DNA samples were from 30 unrelated individuals, who were selected to represent the broad range in age and metabolic parameters of the whole biobank. After removing CpGs for which local genetic polymorphisms could interfere with DNA methylation measurements, and CpGs that did not meet the quality criteria, we recorded methylation of 164 CpG sites, distributed over 104 CpG units, of which 62 contained a single CpG site (Table 2.1). The average methylation of the loci studied ranged from 0 % to 98 % (Figure 2.1). Within the majority of loci, the average methylation of CpG sites was similar. Exceptions were *LEP* with a 34% methylation difference between CpG sites located 18 bases apart and *ABCA1* with a 31% difference between sites 26 bases apart. Previous studies reported methylation differences between men and women [135,157]. We could not detect such differences in an explorative analysis. This may be related to the study size and the number of tests performed.

Table 2.1: Characteristics of methylation assays

Locus	Chromosome	Megabase	Gene function	CpG sites[b]	Single CpG sites[c]	Imprinted	CpG-island	Promoter	Intragenic	Intergenic	Putative mTFBS[d]	Confirmed mTFBS[d]	Transposon	Literature[e]
IL10[a]	01q32.1	205.01	Anti-inflammation	4	2			+			3	6	+	+[136]
NR3C1	05q31.3	142.67	Stress response	20	4		+	+			3	2		+[93]
TNF	06p21.33	31.65	Pro-inflammatory	8	5			+			2			+[138]
IGF2R[a]	06q25.3	160.35	Growth/ Apoptosis	10	0	?	+	+	+		1			+[139]
GRB10	07p12.2	50.82	IIS inhibitor	16	5	+	+	+			4	2		+[140]
LEP[a]	07q32.1	127.67	Metabolism	10	5		+	+				3		+[141]
CRH[a]	08q13.1	67.25	Stress response	5	5			+			1	1		+[142]
ABCA1	09q31.1	106.73	Cholesterol transport	16	3		+	+			4	4		?[143]
IGF2[a]	11p15.5	2.13	Early growth	5	3	+		+	+					+[133]
INSIGF[a]	11p15.5	2.14	(Embryonic) Growth	4	4	?		+			2			+[144]
KCNQ1OT1[a]	11p15.5	2.68	Imprinting control region	14	7	+	+	+						+[145]
MEG3	14q32.2	100.36	Growth suppressor	7	3	+	+	+			1	1		+[146]
FTO	16q12.2	52.38	Development	10	4				+				+	
APOC1[a]	19q13.32	50.11	Metabolism	6	6					+	3			?[147]
GNASAS	20q13.32	56.86	Growth/Lypolytic signal	17	3			+				1		+[148]
GNAS A/B	20q13.32	56.90	Growth/Lypolytic signal	12	3	+	+	+						+[137]

a: With these assays the stability across time and the correlation between tissues was also investigated

b: CpG sites that met the quality criteria described in the methods section

c: CpG sites of which the methylation proportion was measured individually

d: Methylation sensitive transcription factor-binding sites

e: CpG methylation previously reported to associate with gene expression

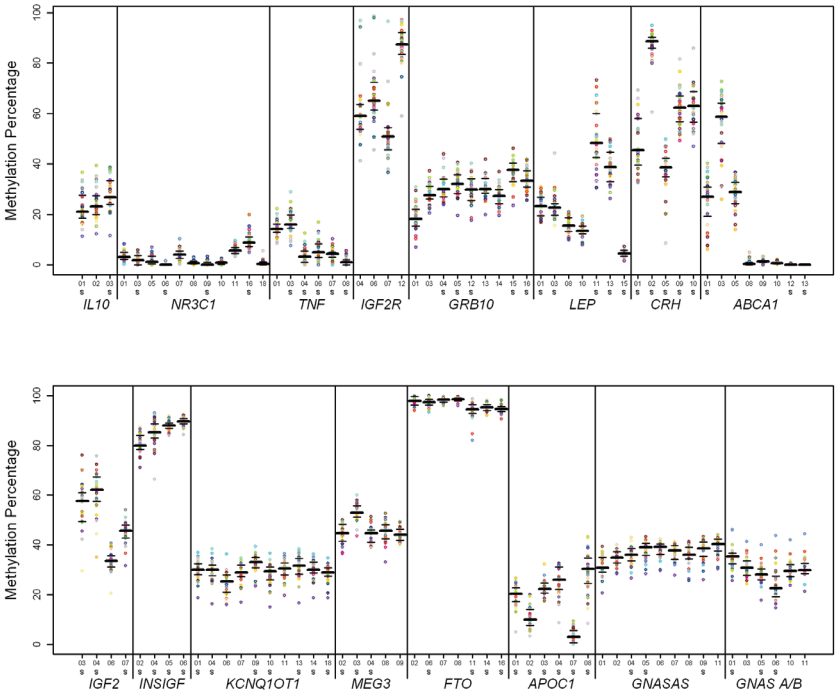


Figure 2.1: Inter-individual variation in DNA methylation
Methylation percentage, y-axis, at every CpG unit, x-axis, for each of the 30 individuals, coloured dots. The order of the loci is based on their chromosomal location, starting at the lowest designation. CpG unit numbers are counted from the Forward primer onwards. The name of each locus is given below the x-axis. The bold horizontal bar gives the median, the thin horizontal bars show the inter-quartile range for each CpG unit. The vertical lines across the plot separate the loci. The corresponding CpG sites of each CpG unit are given in table S2.5B. Individually measured CpG sites are marked with an "S" below the unit number.

A considerable inter-individual variation in CpG methylation was observed. The variation approximated a normal distribution except for CpG units that showed no or very little variation (average DNA methylation close to 0 % or 100 %). The extent of this variation varied per CpG unit (0% <SD<15%). To exclude the possibility that this difference in variation might merely be due to the fact that DNA methylation is truncated at 0 % and 100 %, a variance stabilizing transformation was applied [155]. The variation remained significantly different between CpG units ($p=10^{-16}$). To validate these findings, we measured the DNA methylation of 8 out of 16 loci (55 CpG sites, 31 CpG units with 1 CpG

Table 2.2: Association of neutrophil proportion with DNA methylation

Locus	Variance explained	p-value of effect
IL10	50.1 %	$3.9*10^{-06}$
NR3C1	0.2 %	0.555
TNF	8.0 %	0.037
IGF2R	5.0 %	0.208
GRB10	0.7 %	0.625
LEP	7.4 %	0.019
CRH	4.0 %	0.022
ABCA1	7.3 %	0.021
IGF2	3.5 %	0.185
INSIGF	0.3 %	0.674
KCNQ1OT1	0.4 %	0.733
MEG3	4.1 %	0.165
FTO	0.1 %	0.714
APOC1	6.2 %	0.026
GNASAS	0.8 %	0.590
GNAS A/B	0.0 %	0.889

site) in 34 additional individuals from the NTR biobank, and this yielded similar results (Figure S2.1).

Cellular heterogeneity

DNA methylation was measured on genomic DNA extracted from whole blood. As whole blood consists of different cell types, which may display differences in DNA methylation, we tested to what extent the inter-individual variation observed was influenced by cellular heterogeneity as assessed by counting the major cell subclasses. The largest cellular fraction consisted of neutrophils (average proportion 53.7 %; SD=8.8 %) thus contributing the majority of DNA molecules on which CpG methylation was analysed. For 10 out of 16 loci, the variation in DNA methylation was not associated with this measure of cellular heterogeneity (Table 2.2). For the remaining loci, the variation in DNA methylation that could be explained by variation in cellular heterogeneity was generally small and associations were of borderline significance. *IL10*, which is highly expressed in leukocytes, was a notable exception: 50% of its variation

in DNA methylation could be attributed to the neutrophil cell count. The analysis was repeated using the lymphocyte percentage (average proportion 34.4 %; SD=8.4 %), which was highly correlated with the neutrophil percentage (r = -0.95) and similar results were observed (data not shown). No influence of monocyte percentage was observed on DNA methylation (average 8.0 %; SD=1.8 %; correlation with the neutrophil proportion: r = -0.48). The influence of cellular heterogeneity was not affected by gender (data not shown).

To validate these findings, we performed the same test on the 8 loci in an additional 34 individuals (Table S2.1). The loci previously not showing an association were again not associated with the neutrophil percentage (*IGF2R*, *IGF2*, *INSGF* and *KCNQ1OT1*). Two of the associations of borderline significance were not found (*APOC1* and *CRH*), but the modest association of *LEP* with the neutrophil percentage was replicated ($p=1.0x10^{-4}$). Again, a substantial proportion of the variation in *IL10* methylation could be attributed to the neutrophil percentage (27.9%, $p=8.0x10^{-8}$).

Correlations and patterns of CpG methylation

To investigate patterns of DNA methylation further within and across loci, correlations between the methylation of CpG sites were computed and visualized using a heat map after unsupervised clustering (Figure 2.2). CpG methylation was particularly correlated within loci ($r_{max} = 0.95$) but also across loci ($r_{max} = 0.68$). The cluster of loci correlating irrespectively of chromosomal location included paternally imprinted loci (*MEG3* and *GNASAS*) and maternally imprinted loci (*GRB10, KCNQ1OT1* and *GNAS A/B*). These observations were unaffected by variance-stabilizing transformation or adjustment for cell heterogeneity prior to analysis (data not shown).

The correlation was studied for the same 16 loci in 60 controls of the Dutch Hunger Winter Families Study that we measured previously [14,135]. The heat map of the correlations revealed similar patterns (Figure S2.2). Again, a strong correlation was observed within loci. Across loci, significant correlations were observed for *GNASAS*, *GNAS A/B*, *GRB10* and *MEG3*. In addition, a correlation between these loci and *IGF2* was observed, whereas the correlation with *KCNQ1OT1* was not reproduced. The correlations were similar for both sexes (data not shown).

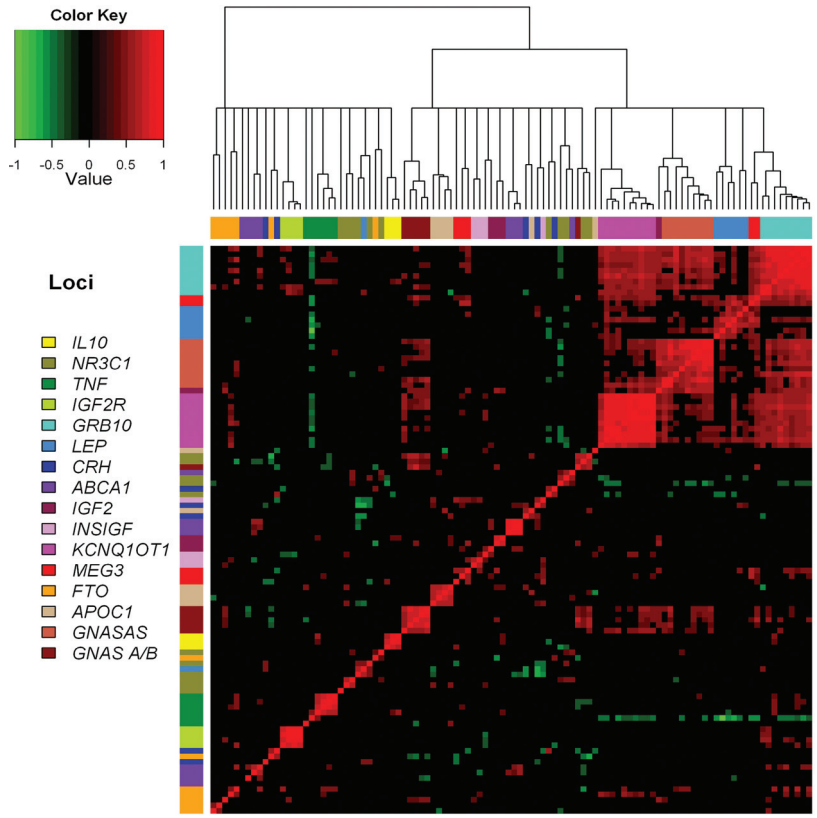


Figure 2.2: Correlation between CpG sites within and across loci
Heat map depicting correlations between the methylation levels of all CpG units of the 16 loci measured in the first group of 30 individuals. For reference the CpG units are annotated by a colour, based on the locus, in the left and upper margin. The diagonal axis running from the lower left to the upper right corner is the line of symmetry where each CpG unit hypothetically correlates with itself. Full correlation (1) is plotted as the brightest red shade, full inverse correlation (-1) is plotted as the brightest green shade, no correlation (0) is plotted as black. Non-significant correlations are depicted as no correlation. The complete clustering is based on the Euclidean distance.

Stability over time

To study the stability of DNA methylation over time, we selected 34 additional individuals from the NTR for whom two blood samples were taken 11-20 years apart. The methylation of 8 loci that were representative of the set of 16 loci was measured (Table 2.1). Overall, DNA methylation was similar at the two time points (Figure 2.3A)

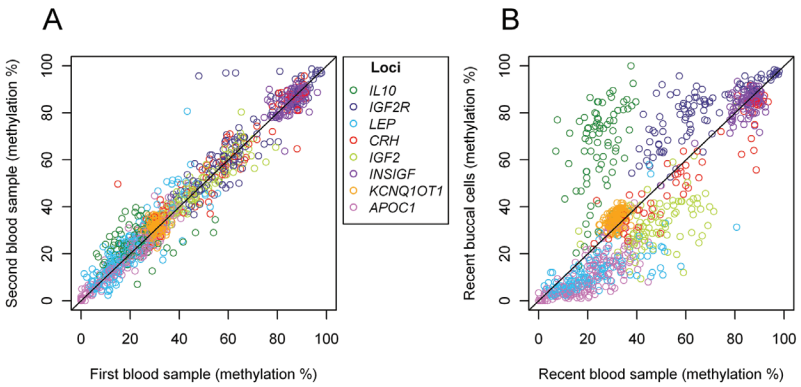


Figure 2.3: Temporal stability and comparison between blood and buccal cell DNA methylation

Scatter plots for individual comparison of CpG methylation between the DNA samples. CpG units of each individual are annotated by colouring based on the locus. The diagonal x = y line is plotted in black for reference. A) CpG methylation in the first blood sample (x-axis) is plotted against methylation in the second, more recent, blood sample (y-axis). Each dot represents one CpG unit of one individual in both DNA samples. B) CpG methylation in the recent blood DNA sample (x-axis) is plotted against methylation in the recent buccal swab DNA sample (y-axis). Each dot represents one CpG unit of one individual in both DNA samples.

and only minor differences were observed (Table 2.3A). Similar average methylation levels between the time-points do not indicate stability per se since methylation may increase in some and decrease in other individuals over time. The variation around the average difference was greatest for *IL10*, which also showed the greatest average difference (-2.8% (SD = 9.1); Table 2.3A). It was lowest for *KCNQ1OT1* (SD = 2.8), indicating relative stability over time. An alternative way to express stability, which takes into account the differences in inter-individual variation of the loci, is to compute correlation coefficients (Table 2.3A). For 5 of the 8 loci, the correlation $\rho > 0.75$ indicated substantial stability between the time points. These loci included *IGF2R* ($\rho = 0.88$) and *APOC1* ($\rho = 0.96$). Note that the correlation was low for *KCNQ1OT1* ($\rho = 0.31$), which can be attributed to the very low level of inter-individual variation. Temporal stability was similar in both sexes (data not shown). From

Table 2.3A: Comparison of DNA methylation in blood samples of the two time points

	Methylation mean (SD)		Difference	Spearman's
Locus	Old blood	New Blood	mean (SD)	rho (ρ)
IL10	22.4% (9.0)	25.2% (6.6)	- 2.8% (9.1)	0.422
IGF2R	65.8% (16.8)	67.6% (17.7)	- 1.8% (8.1)	0.883
LEP	20.0% (11.5)	21.8% (13.0)	- 1.8% (6.1)	0.895
CRH	63.4% (22.1)	63.8% (21.1)	- 0.4% (6.7)	0.942
IGF2	49.4% (11.8)	49.0% (11.3)	0.4% (4.7)	0.924
INSIGF	86.4% (4.1)	85.5% (4.3)	0.9% (3.7)	0.649
KCNQ1OT1	30.8% (2.2)	31.8% (2.6)	- 1.0% (2.8)	0.307
APOC1	19.7% (11.4)	19.6% (11.7)	0.1% (3.4)	0.956

Table 2.3B: Comparison of DNA methylation in recent blood and buccal cell samples

	Methylation, mean (SD)		Difference,	Spearman's
Locus	Blood	Buccal cells	mean (SD)	rho (ρ)
IL10	24.8% (7.1)	64.9% (18.8)	- 40.1% (17.1)	0.442
IGF2R	68.3% (17.4)	81.6% (12.4)	- 13.3% (10.6)	0.827
LEP	21.7% (13.7)	11.8% (8.8)	9.9% (9.9)	0.798
CRH	63.8% (21.9)	58.4% (21.4)	5.4% (7.9)	0.905
IGF2	48.8% (11.5)	32.4% (11.0)	16.4% (10.8)	0.557
INSIGF	85.4% (4.3)	84.4% (4.7)	1.0% (4.6)	0.371
KCNQ1OT1	32.0% (2.6)	34.7% (3.1)	- 2.7% (3.0)	0.481
APOC1	19.9% (11.7)	10.7% (9.6)	9.2% (7.6)	0.822

the same 34 individuals, DNA samples from buccal swabs taken 2-8 years apart were available and showed similar results (Table S2.2).

To exclude the possibility that the higher correlations observed were due to sequence variation not present in dbSNP, we used the mass spectra to identify CpG methylation measurements that were suspected to have been influenced by sequence variation [153]. This was the case for 1 or more individuals for 7/41 CpG units. Removal of these CpG measurements did not affect the correlations $\rho > 0.75$ (*IGF2R*, $\rho = 0.87$; *LEP*, $\rho = 0.90$; *IGF2*, $\rho = 0.92$; *CRH*, $\rho = 0.94$; and *APOC1*, $\rho = 0.95$).

Correlation of CpG methylation between tissues

To test whether DNA methylation in blood could mark that in other tissues, we studied DNA methylation of the 8 loci in the recent blood (mesoderm) and buccal swab (ectoderm) samples of the individuals in whom stability over time was tested. The average level of DNA methylation was generally different between the two tissues and the extent of the difference depended on the locus (Figure 2.3B, Table 2.3B). The variation around the average difference also varied per locus. Again, *IL-10* showed the highest variation (SD = 17.1) and *KCNQ1OT1* the lowest (SD = 3.0). For all loci, the SD of the average difference between the tissues was higher than that of the average difference between the time points. For 4 of the 8 loci, the correlation of DNA methylation between the two tissues was greater than 0.75. These loci included *IGF2R* ($\rho = 0.83$) and *APOC1* ($\rho = 0.82$). Again, the correlations were similar in both sexes (data not shown). After removing CpGs measurements suspected to be influenced by sequence variation not present in dbSNP, similar correlations were found (*IGF2R*, $\rho = 0.82$; *LEP*, $\rho = 0.80$; *CRH*, $\rho = 0.90$; and *APOC1*, $\rho = 0.81$).

Discussion

Epigenetic risk factors are thought to contribute to the development of common diseases such as cardiovascular and metabolic disease [122–124]. Here we investigated whether genomic DNA from existing biobanks is suitable for the identification of these risk factors in epidemiological studies [98,119].

Using genomic DNA from the Netherlands Twin Register biobank [113,114], we first assessed the inter-individual variation in DNA methylation for 16 candidate loci, since the human epigenome map is still in development [41] and epigenome-wide resources on variation (i.e. the epivariome) are lacking. We observed considerable variation in CpG methylation between individuals, except for loci that are either not methylated or fully methylated. The extent of this variation varied between CpG sites. Earlier reports frequently characterized CpG methylation as hypo-, iso- or hypermethylation [133,158]. Our data support our own

previous work [60] and that of others [127] in which DNA methylation was more accurately described as a quantitative trait.

Secondly, we addressed the possibility that the variation in DNA methylation could simply be attributed to cellular heterogeneity in leukocytes between individuals [103]. Blood, like any tissue, consists of a mixture of different cell types that all may have a cell-specific epigenome [24]. Our results show that for the large majority of candidate loci, inter-individual differences in the cellular composition of the blood sample did not contribute to the variation observed in DNA methylation or explained only a minor proportion of this variation. One notable exception was the *IL10* locus for which cellular heterogeneity explained up to half of the total variation in DNA methylation. If cell counts are available for whole blood samples stored in a biobank, the potentially confounding influence of cellular heterogeneity can be monitored using standard statistical methods. If no data on cellular heterogeneity are available, it may be necessary to exclude the association of cellular heterogeneity either with the outcome of interest or with methylation of the locus studied. The latter can be addressed for example in a (sub) study for which data on leukocyte populations are available. Our study suggests that no such relationship will be observed for many loci, in which case biobanks without data on cellular heterogeneity may still be useful.

Thirdly, we investigated patterns in CpG methylation within and across loci. We found that within the locus CpG methylation is highly correlated, except for candidate loci that were not methylated or fully methylated, which corroborates recent findings [60,90]. This observation suggests that assessing the methylation of a subset of CpGs is sufficient to cover the variation in DNA methylation at a locus. This is analogous to genetic association studies in which a small number of tagging SNPs can cover all genetic variation at a locus due to linkage disequilibrium [159]. Moreover, our results provide the first indication that methylation of CpG sites can also be correlated irrespectively of their chromosomal location. This was observed for a subset of mainly imprinted loci, which may be related to the mechanisms responsible for establishing methylation marks at DMRs [160,161].

Fourthly, since DNA methylation is a reversible process [162], it may not be stable over time. If so, this would preclude conclusions about causality in epidemiological studies, since DNA methylation may change during a follow-up period or the differences in DNA methylation may be the consequence of disease [69]. The majority of loci tested were stable over time in DNA from blood and buccal cells despite possible changes in cellular composition during the follow-up period. The fact that we investigated DNA samples that were taken 11 to 20 years apart, implies that these DNA methylation marks may be investigated in most prospective cohort studies in which participants are followed for the development of disease for similar or shorter follow-up periods. However, for a minority of loci we found that although, on average, there was no difference in DNA methylation between the time points, the correlation was lower, indicating relaxed maintenance of these DNA methylation marks. This data resembles recent results on global DNA methylation studying similar follow-up periods [131]. The age of the individuals in our study was limited to young and middle ages (14 to 62 years old). Therefore, we cannot exclude instability over very long periods of time, nor can we exclude the occurrence of greater changes in old age. Indeed, instability of the DNA methylation marks in old age has been reported for both locus-specific [14,132] and global [163] DNA methylation. Moreover, our study did not address the possible occurrence of changes in DNA methylation as a consequence of disease or processes preceding its clinical manifestation [69].

DNA from existing biobanks generally is extracted from easily accessible tissues such as blood. Future studies may reveal DNA methylation patterns in such tissues that mark the risk of disease. As a first step towards establishing a possible causal role, it will be necessary to determine that DNA methylation measured in peripheral tissues is associated with that in tissues directly involved in the disease of interest. Although DNA methylation is thought to be a mechanism driving cell differentiation leading to tissue-specific differentially methylated regions [129], initial reports indicated that DNA methylation measured in blood may be informative. For example, *IGF2* and *ER-α* methylation in blood marked that of colon tissue [133,134]. Also, an autopsy study of 6 subjects and 11 tissues, which did not include

blood, suggested that the hypo- and hypermethylation status of loci is commonly preserved across tissues. Comparing the methylation of candidate loci in blood and buccal cells, we found that for half of the loci tested, DNA methylation measured in blood was a marker for that in buccal cells. These results are promising since blood and buccal cells stem from different germ layers (mesoderm and ectoderm, respectively) and warrant the investigation of correlations with other tissues involved in disease. Genome-scale studies in particular will be informative for defining the (sequence) characteristics of loci showing correlations across tissues. Such studies will be required to interpret the results of epidemiological studies on DNA methylation in blood in a meaningful way.

Our study on the suitability of DNA from existing biobanks for epigenetic studies provides leads for setting up new biobanks specifically aimed at epigenetic epidemiology. Since the correlation between DNA methylation, as measured in DNA from blood and other tissues (directly involved in disease), appears to be complex and locus-dependent, such initiatives should ideally include efforts to sample tissues others than blood (mesoderm), at least for a subgroup representing the cohort. Tissues representing the three germ layers, for which collection is feasible, include: tissue from the mesoderm: biopsies of skeletal muscle, subcutaneous fat and the dermal layer of a skin punch biopsy (fibroblasts); tissue from the ectoderm: the epidermal layer of a skin punch biopsy (keratinocytes) and buccal cells; and tissue from the endoderm: a urine sample (bladder lining) and a stool sample (colonic mucosa). In addition, follow-up sampling of the various tissues should be included (at least for a sub-group) to assess temporal stability and changes in DNA methylation as a consequence of pathology. To account for the cellular heterogeneity of blood samples, blood cell populations should be counted if whole blood is biobanked. This can easily be done using cheap, routine methods. An alternative approach to reduce the cellular heterogeneity is to store peripheral blood mononuclear cells (PBMCs) instead of whole blood. PBMCs include lymphocytes (T and B cells), and monocytes, while the granulocytes (mainly neutrophils) are lost. To completely remove cellular heterogeneity, cells can be separated using magnetic-activated cell sorting (MACS)., However, this is exceedingly costly and will not be feasible for larger numbers in most projects.

Taken together, our results indicate that there are good prospects for the use of existing biobanks for epigenetic studies. Loci that are suitable for testing in epigenetic studies demonstrate inter-individual variation in DNA methylation, stability of this variation in DNA methylation over time and a correlation between DNA methylation as measured blood and the tissue of interest. Our data shows that meeting these criteria is locus-dependent. Therefore, it may be necessary to address this issue for each combination of locus, tissue and disease in new studies.

Acknowledgements

We would like to thank Dr. L.H. Lumey, Mailman School of Public Health, Columbia University, New York, for access to the Dutch Hunger Winter Family Study samples used in our analysis.

Funding

This work was supported by grants from the Netherlands Heart Foundation [NHS2006B083]; the Netherlands Organization for Scientific Research NWO; Twin-family database for behavior genomics studies [480-04-004, 911–03-016]; Spinozapremie [SPI 56-464-14192]; The Centre for Medical Systems Biology (CMSB) in the framework of the Netherlands Genomics Initiative (NGI) and the Centre for Neurogenomics and Cognitive Research (CNCR-VU); Genomewide analyses of European twin and population cohorts [EU/QLRT-2001-01254]; and the European Union–funded Network of Excellence LifeSpan (FP6 036894).

Table S2.1: Association of neutrophil proportion with DNA methylation of the recent blood samples of 34 additional individuals from the NTR biobank

Locus	Variance explained	p-value of effect
IL10	27.9 %	8.0*10-08
IGF2R	1.2 %	0.312
LEP	1.8 %	1.0*10-04
CRH	0.0 %	0.993
IGF2	0.6 %	0.378
INSIGF	0.1 %	0.677
KCNQ1OT1	0.0 %	0.808
APOC1	0.1 %	0.763

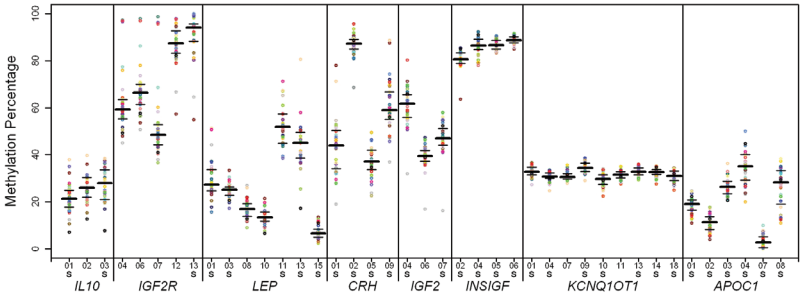


Figure S2.1: Methylation percentage, y-axis, at every CpG-unit, x-axis, for each individual, coloured dots. Measurement was done on DNA from the recent blood samples of the 34 individuals used in the experiments on temporal stability. The order of the loci is based on their chromosomal location, starting at the lowest designation. CpG-unit numbers are counted from the Forward primer onward. The name of each locus is given below the x-axis. The bold horizontal bar gives the median, the thin horizontal bars show the inter-quartile range for each CpG-unit. The vertical lines across the plot separate the loci. The corresponding CpG-sites of each CpG-unit are given in table S2.5C. Individually measured CpG-sites are marked with an “S” below the unit number.

Table S2.2: Comparison of DNA methylation in buccal cell samples from 34 individuals at two time points

	Methylation mean (SD)			
Locus	Old Buccal cells	New Buccal cells	Difference mean (SD)	Spearman's rho (ρ)
IL10	66.9% (13.6)	65.8% (18.4)	1.1% (17.0)	0.613
IGF2R	81.0% (12.9)	82.9% (11.7)	- 1.9% (9.0)	0.701
LEP	10.7% (7.4)	11.0% (7.8)	- 0.3% (7.3)	0.608
CRH	62.7% (21.0)	62.3% (21.5)	0.4% (9.2)	0.888
IGF2	32.5% (8.8)	32.3% (11.0)	0.2% (9.0)	0.686
INSIGF	84.6% (4.4)	84.4% (4.7)	0.2% (3.9)	0.543
KCNQ1OT1	35.8% (3.0)	34.9% (3.0)	0.9% (4.0)	0.097
APOC1	9.4% (7.2)	10.8% (9.7)	- 1.4% (6.1)	0.836

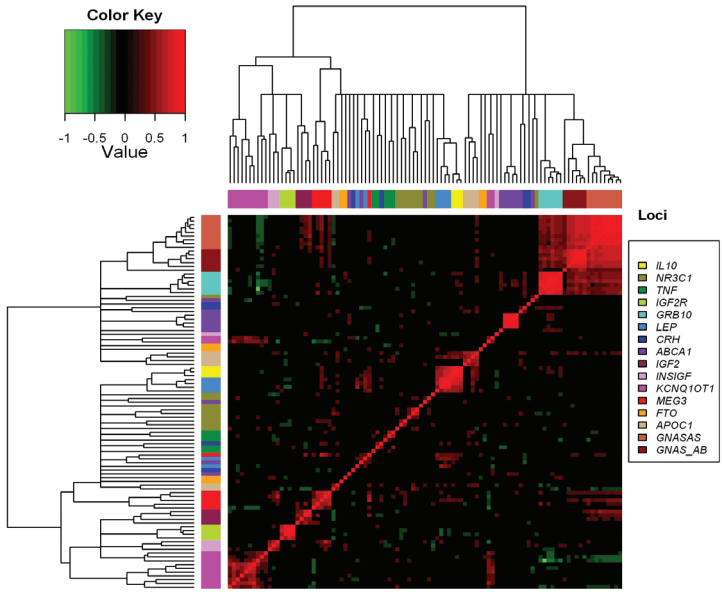


Figure S2.2: Heatmap depicting correlations between the methylation levels of all CpG-units of the 16 loci measured in the controls of the periconceptional group in the Dutch Hunger Winter Families study [14,135] (P.I: Dr. L.H. Lumey, Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA). Correlations were computed on the data after variance stabilizing transformation. For reference the CpG-units are annotated by a colour, based on the locus, in the left and upper margin. The diagonal axis running from the lower left to the upper right corner is the line of symmetry where each CpG-unit hypothetically correlates with itself. Full correlation (1) is plotted as the brightest red shade, full inverse correlation (-1) is plotted as the brightest green shade, no correlation (0) is plotted as black. Non-significant correlations are depicted as no correlation. The complete clustering is based on the Euclidean distance.

Table S2.3: Characteristics of 30 individuals, selected from the NTR biobank, studied

Male, no. (%)	12	(40 %)
Female, no. (%)	18	(60 %)
Smoker, no. (%)	3	(10 %)
Ex-smoker, no. (%)	15	(50 %)
Non-smoker, no. (%)	12	(40 %)
Age at biobanking, mean (SD)	48	(16)
Waist circumference in cm, mean (SD)	98	(22)
LDL-cholesterol (in mmol/L) , mean (SD)	2.47	(1.63)
HDL-cholesterol (in mmol/L) , mean (SD)	1.51	(0.94)
Glucose (in mmol/L) , mean (SD)	9.10	(4.83)

Table S2.4: Age and longitudinal sampling of 34 individuals from the NTR biobank

Age at first sampling (blood)	Years to first follow-up (buccal swab)	Years to second follow-up (blood and buccal swab)
62	6	13
62	6	13
60	9	12
54	6	14
52	6	13
52	6	13
48	6	13
48	6	13
46	10	12
46	10	12
45	10	12
45	10	12
43	6	13
43	6	12
42	6	11
41	6	13
41	10	12
41	6	12
41	10	12
39	6	11
39	9	12
36	10	12
36	10	14
36	10	12
34	9	12
21	13	15
19	10	16
17	15	17
17	14	16
16	11	16
16	13	16
14	17	19
14	16	18
14	15	19

Table S2.5A: Primers used in bisulfite PCR

Locus	Forward primer	Reverse primer
IL10	TGATTGGTTGAATAT-GAATTTTTGTAT	CACCCCCTCATTTTTACTTAAAAA
NR3C1	GATTTGGTTTTTTTGGGG	TCCCTTCCCTAAAACCT
TNF	GGGTATTTTTGATGTTTGT-GTGTT	CAATACTCATAATATCCTTTCCAAAAAA
IGF2R	AGGTAGAAAAAGGTTTTG-GAAG	CAAATCTTAAAAACTAACTAAAAACC
GRB10	GGAATTTTAGGATTAAATT-TATGTGA	AACTTCCAAAAAAAACCTCTCC
LEP	GTTTTTGGAGGGATATTA-AGGATTT	CTACCAAAAAAAACCAACAAAAAAA
CRH	TGGTTGTTGTTTTTTTGG-TAGG	AATTTCTCCACTCCAAAACCTAAA
ABCA1	ATTTTATTGGTGTTTTTG-GTTGT	ATCAAAACCTATACTCTCCCTCCTC
IGF2	TGGATAGGAGATTGAG-GAGAAA	AAACCCCAACAAAAACCACT
INSIGF	GTTTTGAGGAAGAGGT-GTTGA	ACCTAAAATCCAACCACCCTAA
KCNQ1OT1	TTTGGTAGGATTTTGTTGAG-GAGTTTT	CTCACACCCAACCAATACCTCATAC
MEG3	TTTTTTTTAATAGTATTTT-GATTTTTG	AAATAATCCCCACACACATACC
FTO	GTTTGTAATTTTAG-TATTTTGGGAGGT	TTTATTTCCATTTATCCATTCTCAAA
APOC1	GGAGGAGGGAGATTAATAT-TAATTTGT	ACCCCAAACCTATAACCACCTT
GNASAS	GTAATTTGTGGTATGAG-GAAGAGTGA	TAAATAACCCAACTAAATCCCAACA
GNAS A/B	ATGATTTAATTAAGGTTTTAG-GAAAGG	TAAAAATACAAAACCTCCCCTACTC
10-mer tag	AGGAAGAGAG + primer	
T7 tag		CAGTAATACGACTCACTATAGGGAGA-AGGCT + primer

Table S2.5B: CpG-sites per fragment of the loci that were analyzed for variation

Locus	CpG-sites[a] analyzed
IL10	1, 2&3, 4
NR3C1	1&2, 4, 7&8, 9, 10&11, 12&13, 14, 15&16, 17-20, 31, 33&34
TNF	1-3, 5, 6, 9, 10, 11
IGF2R	4&5, 8-10, 11-13, 20&21
GRB10	1&2, 4-6, 7, 8, 17, 18-21, 22&23, 24, 25
LEP	1, 8, 16&17, 19-21, 22, 25, 27
CRH	1, 2, 5, 9, 10
ABCA1	1, 3&4, 6-9, 15&16, 17&18, 19-21, 24, 25
IGF2	3, 4, 6&7, 8
INSIGF	2, 4, 5, 6
KCNQ1OT1	1, 6, 8&9, 10-12, 15, 16, 17&18, 20, 21, 25
MEG3	2, 3, 4, 8&9, 10&11
FTO	2&3, 7, 8&9, 10&11, 14, 17, 19
APOC1	1, 2, 3, 4, 10, 11
GNASAS	1&2, 3&4, 6, 7, 8&9, 10-12, 13&14, 15, 17-19
GNAS A/B	1, 3&4, 7, 8, 13-15, 16-19

a: CpG-site number is counted from the forward primer onward

Table S2.5C: CpG-sites per fragment of the loci that were analyzed for stability

Locus	CpG-sites[a] analyzed
IL10	1, 2&3, 4
IGF2R	4&5, 8-10, 11-13, 20&21, 22[b]
LEP	1, 8, 16&17, 19-21, 22[b], 25, 27
CRH	1[b], 2, 5, 9
IGF2	4, 6&7, 8
INSIGF	2, 4, 5, 6
KCNQ1OT1	1, 6, 10-12, 15, 16, 17&18, 20, 21, 25
APOC1	1, 2, 3, 4, 10, 11

a: CpG-site number is counted from the forward primer onward
b: CpG-site measurement met the quality criteria only in the recent blood samples

Table S2.6: All CpG-units of the 16 loci

Locus_CpG-unit[a]	CpG-sites[a]	Reason for removal prior to quality control
*IL10*_01	CpG_1	
*IL10*_02	CpG_2&3	
*IL10*_03	CpG_4	
*NR3C1*_01	CpG_1&2	
*NR3C1*_02	CpG_3	
*NR3C1*_03	CpG_4	
*NR3C1*_04	CpG_5&6	
*NR3C1*_05	CpG_7&8	
*NR3C1*_06	CpG_9	
*NR3C1*_07	CpG_10&11	
*NR3C1*_08	CpG_12&13	
*NR3C1*_09	CpG_14	
*NR3C1*_10	CpG_15&16	
*NR3C1*_11	CpG_17-20	
*NR3C1*_12	CpG_21	Mass overlap with unit 14
*NR3C1*_13	CpG_22-28	
*NR3C1*_14	CpG_29	Mass overlap with unit 12
*NR3C1*_15	CpG_30	
*NR3C1*_16	CpG_31	
*NR3C1*_17	CpG_32	
*NR3C1*_18	CpG_33&34	
*NR3C1*_19	CpG_35-41	High Mass
*NR3C1*_20	CpG_42	rs5871844 and rs34027900
*TNF*_01	CpG_1-3	
*TNF*_02	CpG_4	
*TNF*_03	CpG_5	
*TNF*_04	CpG_6	
*TNF*_05	CpG_7&8	
*TNF*_06	CpG_9	
*TNF*_07	CpG_10	
*TNF*_08	CpG_11	
*IGF2R*_01	CpG_1	Mass overlap with unit 8
*IGF2R*_02	CpG_2	Mass overlap with unit 3
*IGF2R*_03	CpG_3	Mass overlap with unit 2
*IGF2R*_04	CpG_4&5	

Locus_CpG-unit[a]	CpG-sites[a]	Reason for removal prior to quality control
*IGF2R*_05	CpG_6&7	Mass overlap with unit 11
*IGF2R*_06	CpG_8-10	
*IGF2R*_07	CpG_11-13	
*IGF2R*_08	CpG_14	Mass overlap with unit 1
*IGF2R*_09	CpG_15&16	Mass overlap with unit 10 and rs677882 and rs8191722
*IGF2R*_10	CpG_17	Mass overlap with unit 9
*IGF2R*_11	CpG_18&19	Mass overlap with unit 5 and rs8191721 and rs8191720
*IGF2R*_12	CpG_20&21	
*IGF2R*_13	CpG_22	
*GRB10*_01	CpG_1&2	
*GRB10*_02	CpG_3	Mass overlap with unit 9
*GRB10*_03	CpG_4-6	
*GRB10*_04	CpG_7	
*GRB10*_05	CpG_8	
*GRB10*_06	CpG_9&10	Mass overlap with unit 8
*GRB10*_07	CpG_11	Low mass
*GRB10*_08	CpG_12	Mass overlap with unit 6
*GRB10*_09	CpG_13	Mass overlap with unit 2
*GRB10*_10	CpG_14&15	
*GRB10*_11	CpG_16	
*GRB10*_12	CpG_17	
*GRB10*_13	CpG_18-21	
*GRB10*_14	CpG_22&23	
*GRB10*_15	CpG_24	
*GRB10*_16	CpG_25	
*LEP*_01	CpG_1	
*LEP*_02	CpG_2-7	High Mass and rs791620
*LEP*_03	CpG_8	
*LEP*_04	CpG_9&10	Mass overlap with unit 9
*LEP*_05	CpG_11	Mass overlap with units 12 and 6
*LEP*_06	CpG_12&13	Mass overlap with units 5 and 12
*LEP*_07	CpG_14&15	Mass overlap with unit 14
*LEP*_08	CpG_16&17	
*LEP*_09	CpG_18	Mass overlap with unit 4

Locus_CpG-unit[a]	CpG-sites[a]	Reason for removal prior to quality control
*LEP*_10	CpG_19-21	
*LEP*_11	CpG_22	
*LEP*_12	CpG_23&24	Mass overlap with units 5 and 6
*LEP*_13	CpG_25	
*LEP*_14	CpG_26	Mass overlap with unit 7
*LEP*_15	CpG_27	
*LEP*_16	CpG_28	
*LEP*_17	CpG_29	rs2167270
*LEP*_18	CpG_30-32	High Mass
*CRH*_01	CpG_1	
*CRH*_02	CpG_2	
*CRH*_03	CpG_3	
*CRH*_04	CpG_4	
*CRH*_05	CpG_5	
*CRH*_06	CpG_6	Mass overlap with unit 7
*CRH*_07	CpG_7	Mass overlap with unit 6
*CRH*_08	CpG_8	
*CRH*_09	CpG_9	
*CRH*_10	CpG_10	
*ABCA1*_01	CpG_1	
*ABCA1*_02	CpG_2	
*ABCA1*_03	CpG_3&4	
*ABCA1*_04	CpG_5	
*ABCA1*_05	CpG_6-9	
*ABCA1*_06	CpG_10-13	rs2246298
*ABCA1*_07	CpG_14	
*ABCA1*_08	CpG_15&16	
*ABCA1*_09	CpG_17&18	
*ABCA1*_10	CpG_19-21	
*ABCA1*_11	CpG_22&23	rs13306071
*ABCA1*_12	CpG_24	
*ABCA1*_13	CpG_25	
*ABCA1*_14	CpG_26&27	rs2740483
*IGF2*_01	CpG_1	rs3741208 and rs17883577
*IGF2*_02	CpG_2	rs3741209

Locus_CpG-unit[a]	CpG-sites[a]	Reason for removal prior to quality control
*IGF2*_03	CpG_3	
*IGF2*_04	CpG_4	
*IGF2*_05	CpG_5	rs4930041
*IGF2*_06	CpG_6&7	
*IGF2*_07	CpG_8	
*INSIGF*_01	CpG_1	Low mass
*INSIGF*_02	CpG_2	
*INSIGF*_03	CpG_3	
*INSIGF*_04	CpG_4	
*INSIGF*_05	CpG_5	
*INSIGF*_06	CpG_6	
*KCNQ1OT1*_01	CpG_1	
*KCNQ1OT1*_02	CpG_2	
*KCNQ1OT1*_03	CpG_3-5	Mass overlap with unit 8
*KCNQ1OT1*_04	CpG_6	
*KCNQ1OT1*_05	CpG_7	Mass overlap with units 15 and 19
KCNQ1OT1_06	CpG_8&9	
KCNQ1OT1_07	CpG_10-12	
KCNQ1OT1_08	CpG_13&14	Mass overlap with unit 3
KCNQ1OT1_09	CpG_15	
KCNQ1OT1_10	CpG_16	
KCNQ1OT1_11	CpG_17&18	
KCNQ1OT1_12	CpG_19	
KCNQ1OT1_13	CpG_20	
KCNQ1OT1_14	CpG_21	
KCNQ1OT1_15	CpG_22	Mass overlap with units 5 and 19
KCNQ1OT1_16	CpG_23	rs7940500
KCNQ1OT1_17	CpG_24	rs379976
KCNQ1OT1_18	CpG_25	
KCNQ1OT1_19	CpG_26&27	Mass overlap with units 5 and 15
*MEG3*_01	CpG_1	
*MEG3*_02	CpG_2	
*MEG3*_03	CpG_3	
*MEG3*_04	CpG_4	
*MEG3*_05	CpG_5	Mass overlap with unit 6

Locus_CpG-unit[a]	CpG-sites[a]	Reason for removal prior to quality control
*MEG3*_06	CpG_6	Mass overlap with unit 5
*MEG3*_07	CpG_7	
*MEG3*_08	CpG_8&9	
*MEG3*_09	CpG_10&11	
*MEG3*_10	CpG_12-14	High Mass
*FTO*_01	CpG_1	Mass overlap with Unit 5
*FTO*_02	CpG_2&3	
*FTO*_03	CpG_4	Low mass
*FTO*_04	CpG_5	
*FTO*_05	CpG_6	Mass overlap with Unit 1
*FTO*_06	CpG_7	
*FTO*_07	CpG_8&9	
*FTO*_08	CpG_10&11	
*FTO*_09	CpG_12	Mass overlap with Unit 17
*FTO*_10	CpG_13	
*FTO*_11	CpG_14	
*FTO*_12	CpG_15	
*FTO*_13	CpG_16	
*FTO*_14	CpG_17	
*FTO*_15	CpG_18	
*FTO*_16	CpG_19	
*FTO*_17	CpG_20	Mass overlap with Unit 9
*APOCI*_01	CpG_1	
*APOCI*_02	CpG_2	
*APOCI*_03	CpG_3	
*APOCI*_04	CpG_4	
*APOCI*_05	CpG_5&6	rs402204
*APOCI*_06	CpG_7-9	High Mass and rs5111
*APOCI*_07	CpG_10	
*APOCI*_08	CpG_11	
*GNASAS*_01	CpG_1&2	
*GNASAS*_02	CpG_3&4	
*GNASAS*_03	CpG_5	
*GNASAS*_04	CpG_6	
*GNASAS*_05	CpG_7	

Locus_CpG-unit[a]	CpG-sites[a]	Reason for removal prior to quality control
*GNASAS*_06	CpG_8&9	
*GNASAS*_07	CpG_10-12	
*GNASAS*_08	CpG_13&14	
*GNASAS*_09	CpG_15	
*GNASAS*_10	CpG_16	rs45596642
*GNASAS*_11	CpG_17-19	
*GNAS A/B*_01	CpG_1	
*GNAS A/B*_02	CpG_2	Mass overlap with unit 4
*GNAS A/B*_03	CpG_3&4	
*GNAS A/B*_04	CpG_5&6	Mass overlap with unit 2
*GNAS A/B*_05	CpG_7	
*GNAS A/B*_06	CpG_8	
*GNAS A/B*_07	CpG_9&10	
*GNAS A/B*_08	CpG_11	Low mass
*GNAS A/B*_09	CpG_12	
*GNAS A/B*_10	CpG_13-15	
*GNAS A/B*_11	CpG_16-19	
Totals		
# of amplicons	16	
total # units	191	
# units outside detection range	9	
# units with equal or overlapping mass	36	
# units with potential SNP	12	
Total # CpG-units removed	57	

a: CpG-unit and CpG-site numbers are counted from the forward primer onward

Chapter 3

DNA Methylation differences after exposure to prenatal famine are common and timing- and sex-specific

Elmar W. Tobi[1], L.H. Lumey[3,5], Rudolf P. Talens[1], Dennis Kremer[1], Hein Putter[2], Aryeh D. Stein[4], P. Eline Slagboom[1], Bastiaan T. Heijmans[1,*]

1. Molecular Epidemiology, Leiden University Medical Center, Leiden, 2300RC, The Netherlands
2. Medical Statistics, Leiden University Medical Center, Leiden, 2300RC, The Netherlands
3. Endocrinology, Leiden University Medical Center, Leiden, 2300RC, The Netherlands
4. Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA
5. Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA

Published in Human Molecular Genetics, Volume 18, Issue 21, pages 4046 – 4053, 2009

Abstract

Prenatal famine in humans has been associated with various later-life consequences depending on the gestational timing of the insult and the sex of the exposed individual. Epigenetic mechanisms have been proposed to underlie these associations. Indeed, animal studies and our early human data on the imprinted *IGF2* locus indicated a link between prenatal nutritional and DNA methylation. However, it remains unclear how common changes in DNA methylation are and whether they are sex- and timing-specific paralleling the later-life consequences of prenatal famine exposure. To this end we investigated the methylation of 15 loci implicated in growth and metabolic disease in individuals who were prenatally exposed to a war-time famine in 1944-45. Methylation of *INSIGF* was lower among individuals who were periconceptionally exposed to the famine (n=60) as compared with their unexposed same-sex siblings ($P=2x10^{-5}$), whereas methylation of *IL10*, *LEP*, *ABCA1, GNASAS* and *MEG3* were higher (all $P<10^{-3}$). A significant interaction with sex was observed for *INSIGF*, *LEP* and *GNASAS*. Next, methylation of 8 representative loci was compared between 62 individuals exposed late in gestation and their unexposed siblings. Methylation was different for *GNASAS* ($P=1.1x10^{-7}$) and, in men, *LEP* (P=0.017). Our data indicate that persistent changes in DNA methylation may be a common consequence of prenatal famine exposure and that these changes depend on the sex of the exposed individual and the gestational timing of the exposure.

Introduction

Adverse environmental conditions during specific windows of mammalian development can have lasting effects on metabolic pathways and physiology, thereby influencing the susceptibility to chronic diseases [164]. An extensive epidemiologic literature has reported associations between characteristics of early development and health outcomes later in life [165,166]. Historical famines provide a quasi-experimental setting in which the long-term consequences of adverse conditions during development can be studied in humans. Studies of the Dutch Hunger Winter, a severe wartime famine at the end of WWII affecting the western part of The Netherlands, suggest that famine exposure *in utero* can lead to various adverse metabolic or mental phenotypes, depending on the sex of the exposed individual and the timing of the exposure during gestation [167–170].

The period around conception may be especially sensitive to famine exposure [171]. Exposure to famine in this period is associated with diverse phenotypic outcomes such as an increased risk of adult schizophrenia [170] and spina bifida at birth in men [172]. However, the effects of prenatal famine exposure are not limited to this developmental period. An example of a general, sex-specific late life effect is the increase in body mass index [169,173] and various lipids in blood [167] among famine-exposed women, irrespective of the precise gestational timing of the exposure. Also, increases in cerebro-cardiovascular related deaths have been reported among individuals exposed to seasonal food shortages independent of the gestational timing in a historical cohort [174], although preliminary results from the Dutch Famine indicated that an increased risk of coronary artery disease is specific for exposure to famine early in gestation [175]. Animal experiments confirm the importance of timing and sex [176,177]. Thus, prenatal exposure to famine can have different long-term effects that depend on the timing of the exposure and the sex of the exposed individual.

Persistent epigenetic changes induced by environmental factors are a plausible molecular mechanism underlying the relationship between early development and later life disease [119,178]. Experiments in animal models provide strong supporting evidence. Manipulation of the maternal diet during pregnancy lead to a persistent shift in average

DNA methylation levels of specific genes in offspring resulting in permanent changes in coat color or tail shape [128,179]. A proof-of-principle for complex diseases was reported by Bogdarina *et al*., who showed in rats that a low protein diet during pregnancy was associated with decreased DNA methylation of the *Agtr1b* gene promoter in offspring, explaining the increase in blood pressure among these animals [13]. We recently showed that similar mechanisms may be operative in humans; periconceptional exposure to famine was associated with a decrease in DNA methylation of the insulin-like growth factor 2 (*IGF2*) differentially methylated region (DMR) [14]. However, to be a valid candidate mechanism in humans, the effect of prenatal famine on DNA methylation should be widespread, mirror the epidemiologic findings with sex- and timing-specific associations and affect genes in relevant pathways.

To further explore associations between prenatal famine and DNA methylation, including the role of the timing of exposure and the sex of the prenatally exposed individual, we assessed the methylation state of loci in 15 candidates genes involved in metabolic and cardiovascular disease and growth with diverse epigenetic features in our ongoing Hunger Winter Families Study [11].

Results

Periconceptional exposure

We studied methylation of loci implicated in the transcriptional regulation of 15 candidate genes for metabolic and cardiovascular disease. The loci studied included imprinted loci (*GNASAS*, *GNAS A/B*, *MEG3*, *KCNQ1OT1, INSIGF* and *GRB10*), a putatively imprinted locus (*IGF2R*) and non-imprinted loci (*IL10*, *TNF*, *ABCA1*, *APOC1*, *FTO*, *LEP*, *NR3C1* and *CRH*). The selection of the loci measured was based on a combination of factors, including binding of methylation sensitive transcription factors, associations of DNA methylation with gene expression and the presence of differentially methylated regions for imprinted loci (a detailed overview can be found in table S3.1).

DNA methylation of these 15 loci was measured in 60 individuals conceived during the Dutch Famine (i.e. exposed periconceptionally) and compared with their unexposed same-sex sibling to minimize the possible confounding effects

of familial environment and genetic background. For six of the fifteen loci, significant differences in DNA methylation were observed (Table 3.1). DNA methylation was increased among famine exposed individuals for the imprinted genes *GNASAS* ($P=3.1x10^{-6}$) and *MEG3* ($P=8.0x10^{-3}$) and the non-imprinted *IL10* ($P=1.8x10^{-6}$), *ABCA1* ($P=8.2x10^{-4}$) and *LEP* ($P=2.9x10^{-3}$) proximal promoters. DNA methylation was decreased for the imprinted *INSIGF* promoter ($P=2.3x10^{-5}$), which is part of the proximal promoter of *INS* [180]. For the remaining 9 loci there was no association with periconceptional exposure to famine. All associations remained statistically significant after Bonferroni correction for multiple testing (15 loci) with the exception of *MEG3* ($P_{Bonferroni}= 0.12$). When analysed separately, individual CpG dinucleotides showed similar associations with famine exposure as the complete loci (Figure 3.1 A). The loci affected did not share obvious features with respect to sequence, epigenetic features or biological function.

Late gestational exposure

Epigenetic modulation may also occur during other developmental windows [93,105]. We therefore also studied 62 individuals who were exposed to famine late in gestation together with their unexposed, same-sex siblings. We measured DNA methylation at four loci that were associated with periconceptional famine exposure (*IL10*, *GNASAS*, *INSIGF* and *LEP*) and at four that were not (*IGF2R*, *APOC1*, *KCNQ1OT1* and *CRH*). These loci include four imprinted ones and diverse epigenetic features (Table S3.1).

No associations were observed except for a significant reduction in methylation at the *GNASAS* locus ($P=1.1x10^{-7}$, $P_{Bonferroni}= 8.8 x10^{-7}$) (Table 3.2). This association was consistent for the individual CpG sites within the locus (data not shown). The direction of the association was opposite to what was observed for periconceptional exposure. We then combined all periconceptional and late pregnancy exposed individuals and their controls in a single analysis to test for a statistical interaction between the famine associations with DNA methylation and the precise gestational timing of the exposure. The DNA methylation differences found for *IL10*, *GNASAS* and *INSIGF* were timing-specific (Table 3.2), but for *LEP* the test for interaction was not significant.

Table 3.1. DNA methylation and periconceptional exposure to famine.

Gene locus[a]	Average methylation % (SD)		Within pair difference (Δ%)[b]	Effect size (SD units)[c]	P-value[d]	P-value Bonferroni corrected[e]
IL10	20.8	(6.5)	2.4	0.37	$1.8x10^{-6}$	$2.7x10^{-5}$
GNASAS	48.8	(4.7)	1.1	0.24	$3.1x10^{-6}$	$4.7x10^{-5}$
INSIGF	84.8	(2.6)	-1.6	-0.61	$2.3x10^{-5}$	$3.5x10^{-4}$
LEP	28.6	(4.9)	1.2	0.24	$2.9x10^{-3}$	$4.4x10^{-2}$
MEG3	54.0	(2.4)	0.5	0.21	$8.0x10^{-3}$	0.12
ABCA1	19.9	(4.2)	0.7	0.17	0.017	0.26
ABCA1 meth[f]	36.9	(8.2)	1.7	0.21	$8.2x10^{-4}$	0.012
KCNQ1OT1	30.1	(1.5)	-0.2	-0.16	0.053	n.s.
GRB10	47.2	(4.6)	0.4	0.08	0.091	n.s.
GNASAB	40.3	(5.0)	0.6	0.11	0.092	n.s.
APOC1	16.7	(3.1)	-0.5	-0.17	0.13	n.s.
IGF2R	84.1	(6.9)	-0.7	-0.10	0.29	n.s.
FTO	97.3	(0.8)	-0.5	-0.61	0.28	n.s.
CRH	58.9	(6.0)	0.4	0.07	0.61	n.s.
TNF	9.6	(1.7)	0.1	0.06	0.63	n.s.
NR3C1	4.8	(1.1)	0.0	-0.01	0.79	n.s.

a: Table sorted on P-value.

b: Average absolute difference in DNA methylation between exposed and unexposed siblings.

c: Observed within pair difference divided by the standard deviation in the sibling controls.

d: Two-sided P-value resulting from a linear mixed model accounting for family relations, bisulfite batch and age at blood draw.

e: The Bonferroni corrected P-values. Results that were already not significant before Bonferroni correction are shown as "n.s." (non significant).

f: Separate analysis of methylated CpGs in *ABCA1*. In contrast to the 3' CpGs, 5' CpG dinucleotides (n=13) were methylated. The methylation of the methylated 5' CpGs was highly correlated (R >0.87) amongst themselves, but not with methylation of the 3' CpG dinucleotides, which had little to no methylation (<4.5%).

Sex-specific associations

Previous work found basal DNA methylation differences in humans [157]. We therefore first tested for sex differences in DNA methylation for the measured loci combining the two sibling control groups (men N=56, women N=66). DNA methylation was higher in men for *IGF2R* (+2.6%, P=0.019) and lower in men for *LEP* (-2.6%, $P=3.0x10^{-3}$), *IL10* (-2.9%, P=0.015), and *APOC1* (-1.5, P=0.015) as compared to

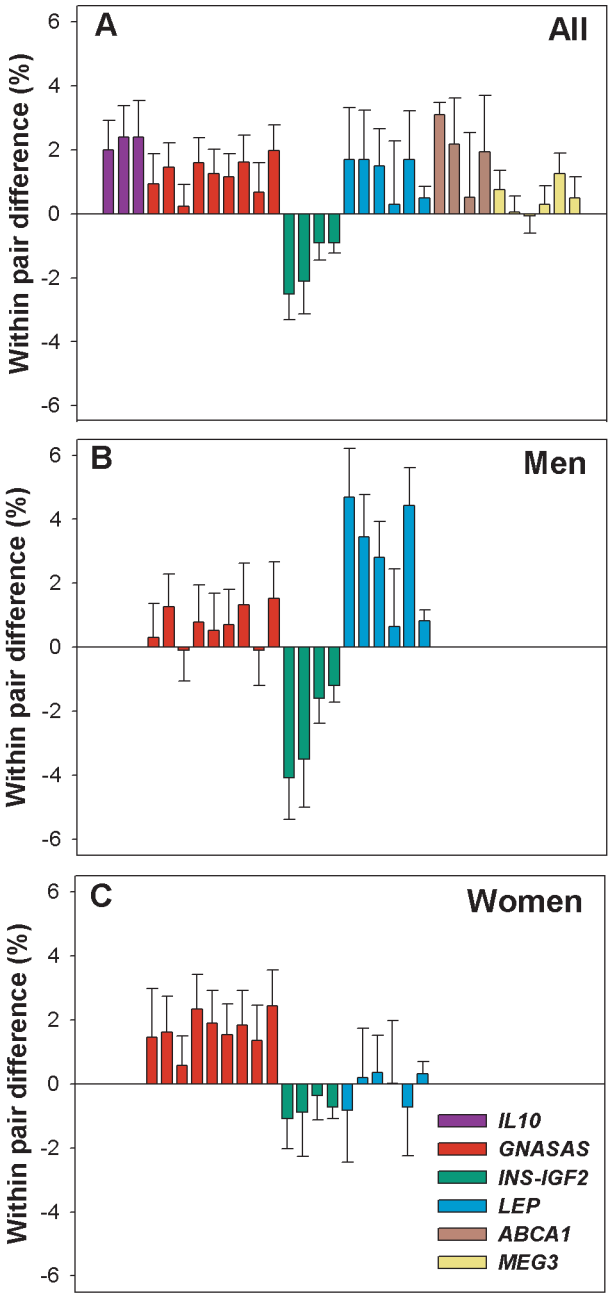


Figure 3.1. Difference in DNA methylation of CpG dinucleotides in siblings discordant for periconceptional exposure to famine.
Bars in the figures represent the average absolute within pair difference in DNA methylation and their standard errors for CpG dinucleotides. A positive difference indicates a higher methylation level among exposed individuals. The exact location of the CpG dinucleotides can be found using supplemental table S3.1. **A**. The absolute within pair difference for CpG dinucleotides for which a significant overall difference of the locus in DNA methylation was observed. **B**. As in panel A but for men. Only loci showing a significant interaction between sex and exposure are depicted. **C**. As in panel B, but for women only.

Table 3.2. DNA methylation and late gestational exposure to famine.

Gene locus[a]	Average methylation % (SD)[b]		Within pair difference (Δ%)[c]	Effect size (SD units)[d]	P-value[e]	P-value Bonferroni corrected[f]	P-value[g] timing specificity
IL10	20.7	(5.0)	-0.2	-0.04	0.76	n.s.	$1.2x10^{-3}$
GNASAS	48.8	(4.2)	-1.1	-0.26	$1.1x10^{-7}$	$8.8x10^{-7}$	$3.1x10^{-12}$
INSIGF	84.7	(2.8)	0.0	0.0	0.95	n.s.	$3.2x10^{-4}$
LEP	28.7	(4.6)	0.4	0.09	0.18	n.s.	0.13
KCNQ1OT1	30.2	(1.7)	0.2	0.12	0.17	n.s.	0.058
APOC1	16.7	(3.1)	-0.6	-0.19	0.22	n.s.	0.90
IGF2R	84.0	(4.4)	0.0	0.0	0.88	n.s.	0.25
CRH	58.9	(4.8)	0.5	0.10	0.51	n.s.	0.86

a: This table has been given the same order as table 3.1.

b: The batch corrected average methylation for the unexposed late gestation sibling controls and the standard deviation.

c: Average absolute difference in DNA methylation between exposed and unexposed siblings.

d: Observed within pair difference divided by the standard deviation in the sibling controls.

e: Two-sided P-value resulting from a linear mixed model accounting for family relations, bisulfite batch and age at blood draw. This test was performed on the late gestational exposed sibships (N=62)

f: The Bonferroni corrected P-values. Results that were already not significant before Bonferroni correction are shown as "n.s." (non significant).

g: Two-sided P value resulting from the test for timing specificity. The timing specificity was calculated by joining the datasets for both the periconceptional and late gestational siblings and their unexposed same-sex siblings and introducing an interaction term for gestational timing times exposure status. This test thus includes all 122 pairs.

women. No significant differences were found for the other 4 loci (*GNASAS*, *INSIGF*, *KCNQ1OT1* and *CRH*).

Next, we tested if the observed significant associations with prenatal famine were sex-specific. The interaction between sex and periconceptional famine exposure was significant for *LEP* ($P=2.3x10^{-4}$), *INSIGF* ($P=8.5x10^{-3}$) and *GNASAS* (P=0.027) (Table 3.3). For *LEP* and *INSIGF* the association of famine exposure with DNA methylation was restricted to men ($P_{LEP, ♂}=3.6x10^{-7}$, $P_{INSIGF, ♂}=6.5x10^{-6}$) (Figure 3.1 B and 3.1 C). For *GNASAS* the association was significant in both sexes but most pronounced in women ($P_{men}=0.013$; $P_{women}=1.1x10^{-5}$). The association between late gestational famine exposure and *GNASAS* methylation was independent of sex.

Table 3.3. Sex-specific associations of DNA methylation with periconceptional exposure to famine.

Gene locus	$P_{sex\ interaction}$ [a]	Sex	Within pair difference (Δ%)[b]	Effect size (SD units)[c]	P-value[d]
GNASAS	0.027	♂	0.7	0.15	0.013
		♀	1.5	0.33	$1.1x10^{-5}$
INSIGF	$8.5x10^{-3}$	♂	-2.6	-0.99	$6.5x10^{-6}$
		♀	-0.8	-0.29	0.11
LEP	$2.3x10^{-4}$	♂	2.8	0.57	$3.6x10^{-7}$
		♀	-0.2	-0.04	0.70

a: The two sided P-value from the test for sex-specificity of the observed periconceptional effect of prenatal famine. This was tested by entering an interaction term of sex times the exposure status in the linear mixed model.

b: Average absolute difference in DNA methylation between exposed and unexposed siblings.

c: Observed within pair difference divided by the standard deviation in the sibling controls.

d: Two-sided P-value resulting from a linear mixed model accounting for family relations, bisulfite batch and age at blood draw.

Timing independent association

For *LEP* there was no indication for a significant interaction between the famine association with DNA methylation and the gestational timing of the exposure, even though the methylation difference was significant only following periconceptional famine exposure. Since, this association was later found to be male-specific, we tested for an interaction with sex in the late exposure group. Indeed, a significantly higher *LEP* methylation was found for men exposed late in gestation (P=0.017). Analysis of the whole cohort (N=244) including both exposure groups, revealed a significant association between prenatal exposure to famine irrespective of the precise gestational timing (P=0.003). Further analysis suggested that this association was male-specific ($P_{interaction}$= $1.3x10^{-6}$; men: +2.2% (0.83SD), $P=7.5x10^{-8}$; women: P=0.47).

Discussion

We studied the DNA methylation levels of 15 loci for their association with prenatal exposure to the Dutch Famine at the end of WWII. For six of the loci studied, we observed significant differences in DNA methylation after famine exposure during periconception (*INSIGF*, *GNASAS*, *MEG3, IL10*, *LEP* and *ABCA1*). This association differed by sex for three loci (*INSIGF*, *GNASAS* and *LEP*). Of the eight loci tested, exposure to famine late in gestation was associated with methylation for *GNASAS* and for *LEP,* which was specific for men. Of interest, the differences in DNA methylation included both increases and decreases, in one case even at the same locus after exposure during different gestational periods. Together with our previous finding that the *IGF2* DMR is associated with periconceptional exposure [14] our data indicate that an adverse prenatal environment may trigger widespread and persistent changes in DNA methylation.

Our current and previous [14] observations suggest that the periconceptional period may be an especially sensitive exposure period in humans. This might be inherent to mammalian development [181,182] and this hypothesis is also supported by detailed animal studies [183,184]. The association of *GNASAS* and *LEP* with late gestational exposure, however, suggests that environmentally induced DNA methylation changes may not be limited to the periconceptional period. This is in line with findings that methylation of the glucocorticoid receptor promoter depends on postnatal circumstances (e.g., maternal care in rats [93] and child abuse in humans [105]). In addition to timing-specific associations, we observed sex-specific associations for three of the six loci for which DNA methylation was significantly associated with prenatal exposure to famine. In men, *LEP* methylation was associated with prenatal famine irrespective of the timing of exposure. Our observation that the methylation changes in relation to the prenatal environment may be sex-specific is in agreement with the sex-specific methylation changes found in offspring of sheep that were folate and vitamin B_{12} restricted during periconception [184]. How such sex-specific associations can arise is currently unknown, but interactions between sex hormones and the expression of DNA methyltransferases may be a factor [185].

The differences in DNA methylation observed here are comparable, although slightly smaller, than we previously found for *IGF2* (absolute difference of 2.7% (5.2% relative to the mean DNA methylation level in the population) [14]. The smaller average differences may be related to the inherent stochastic nature of epigenetic processes [186] leading to a large variability in responses. The stochastic nature is strikingly illustrated by the large variation observed in the response of *agouti* gene methylation on maternal methyl donor supplementation even though the mice are inbred and the environmental conditions are highly controlled [128]. In human studies, genetic and environmental heterogeneity may further obscure the full impact of prenatal famine. In our study, we tried to minimize the heterogeneity caused by these factors by comparing exposed individuals with their unexposed, same-sex siblings. Another potential source of heterogeneity is the cellular diversity of whole blood, the tissue we studied. This heterogeneity is less likely to play a role for the affected imprinted loci *IGF2* [14], *GNASAS* and *MEG3*, since methylation is generally cell-type independent for imprinted loci [31] and the observed differences for non-imprinted loci between exposed individuals and controls were similar. With respect to our finding that *GNASAS* methylation was associated with exposure to famine late in gestation, it should be mentioned that a compromised late gestational development was hypothesized to lead to an immature immune system. The absolute numbers of lymphocytes were reported to be lower in children with a shorter gestation in a large, population-based study [187]. However, since the proportions of the different cell types were not affected, it is unlikely that this has contributed to our findings.

The changes observed are comparable to those found in the liver of rats prenatally exposed to a protein-deficient diet, where promoter methylation of the *Pparα* promoter was decreased from 6.1% to 4.5% with individual CpG dinucleotides affected up to 5% and explaining up to 43% of the variance in gene expression of *Pparα* [188]. This small absolute decrease of *Pparα* DNA methylation corresponded to a large ~26% relative change. It may be hypothesized that modest absolute changes in DNA methylation may lead to significant changes in gene expression for loci with a relatively low methylation level so that relative changes are substantial. The larger relative changes of *LEP* in men

(2.8/27.1=10.3%) and *IL10* (2.4/20.8=11.5%) may be more likely to have functional consequences than that for *INSIGF* in men (2.6/84.8=3.1%).

Similar to animal studies of methionine restriction [184], we observed both increases and decreases in DNA methylation, depending on the locus studied. Our results cannot be readily explained by damage due to a deficiency in dietary methyl donors due to the famine and may thus be part of an adaptive response. To definitely prove or disprove the existence of an adaptive response it will be necessary to characterize epigenetic responses of entire relevant pathways. Epigenetic mechanisms may contribute to the development of a thrifty phenotype [164,178] and sex differences in this respect are increasingly well described [176]. It has been hypothesized that a thrifty phenotype may result from the combined effects of smaller epigenetic changes across the genome shifting metabolic networks [189]. *INS* [190] and *LEP* [189], both affected by prenatal famine exposure, were suggested to be particular relevant in this respect. Our observations provide empirical evidence for these hypotheses in humans.

Our finding that the association of DNA methylation with prenatal conditions may depend on timing and sex matches the specificity in phenotypic outcomes observed for prenatal famine exposure, including neonatal outcomes [172] and psychiatric [170] and metabolic traits [167,169,173]. It remains to be determined if the observed differences in DNA methylation in blood mark functional differences in relevant tissues for these traits. Previous work on *IGF2* DMR has shown that methylation in blood can mark the methylation level in a relevant tissue [133]. Also, the CpG sites studied in the *LEP* promoter were previously reported to show similar methylation in peripheral blood and adipocytes *in vivo* [191] and to influence *LEP* expression *in vitro* [141]. Changes in DNA methylation induced early in development, like by prenatal famine, may be propagated soma-wide contributing to the observed correlation in DNA methylation across tissues [128,183]. A preliminary analysis of our data did not reveal significant associations between DNA methylation and plasma lipids and BMI that we previously reported to depend on prenatal exposure to famine in women in this cohort [167,169]. But DNA methylation at most genes relevant to these particular complex phenotypes was found

to be affected in men only, and not women. Similar to genetic studies, epigenetic studies linking variation in DNA methylation and complex disease will most likely require large study samples. Such studies may be particularly promising for *IL10,* which we found to be sensitive to periconceptional famine exposure. Genetic variation influencing *IL10* expression at this locus is associated with schizophrenia [192,193], a phenotype also particular to the periconceptional period [170].

In summary, our study shows that exposure to famine in pregnancy may cause persistent changes in DNA methylation levels of multiple imprinted and non-imprinted genes with diverse biological functions. Our data support the hypothesis that associations between early developmental conditions and health outcomes later in life may be mediated by changes in the epigenetic information layer. Understanding how disturbances early in human development are linked to later life disease may suggest new ways to prevent disease.

Methods and materials

Study population

Study participants are a subset of the population from our ongoing Hunger Winter Families Study, whose design and recruitment was described previously [11]. In short, this study includes individuals exposed to famine during different periods in pregnancy and time-controls born before or after the famine. Study subjects were selected from births between 1943-1947 at three institutions in famine-exposed cities (the midwifery training schools in Amsterdam and Rotterdam and the Leiden University Medical Center). The daily rations distributed by the authorities during the famine period from November 28, 1944 to May 15, 1945 period contained an average energy equivalent of 667 kcal (SD, 151). During the famine there was little variation in the percentage of calories from proteins (12%), fat (19%) and carbohydrates (69%) [194]. In addition to time controls from these three institutions, we recruited whenever possible a same-sex unexposed sibling of these individuals to serve as sibling control. Clinical examinations, including blood collections, were completed for 311 births in these institutions and 311 sibling controls. Ethical approval for the study was obtained

from the participating institutions, and all participants provided written informed consent.

The subset selected for the present study includes two exposure groups with their unexposed sibling controls. The first exposure group comprises births conceived during the height of the famine, with periconceptional and early pregnancy exposure. The second exposure group comprises births during the height of the famine, with late pregnancy exposure. Periconceptionally exposed individuals were defined as births with a mother's last menstrual period between November 28, 1944 and May 15, 1945. This group includes 60 individuals (age 58.1y, SD 0.35), of whom 28 were male. Individuals exposed in late pregnancy were defined as births between January 28 and May 30, 1945. This group includes 62 individuals (age 58.8y, SD 0.40), of whom 28 were male. As controls for each exposure group we used a matched same-sex sibling (age 57.1y, SD 5.50). The total study population therefore includes 244 individuals.

DNA methylation measurements

Genomic DNA was isolated from whole blood using the salting out method. One microgram of genomic DNA was bisulfite treated using the EZ 96-DNA methylation kit (Zymo Research). All samples were bisulfite treated on a total of three 96-well plates. Sibling pairs were on the same plate and periconceptionally and last trimester exposed pairs were equally distributed over the plates. DNA methylation of CpG dinucleotides were measured by a mass spectrometry based method (Epityper, Sequenom). The method determines the amount of DNA methylation by interrogating thousands of DNA copies assuming that 1 ng of genomic DNA equals ~ 300 copies. The quantitative nature, accuracy and reproducibility of this method has been shown extensively [14,87,151]. All biochemical steps inherent to the methodology were performed according to the manufacturers' protocol. Bisulfite converted DNA specific PCR primers used to amplify the 15 investigated regions are summarized in the supplementary table S3.1. This table includes the precise genomic location of the amplified regions and an overview of the CpG sites quantized in each amplicon. All individuals exposed early in gestation and their sibling controls were measured in triplicate on a single 384 well plate for each locus and the same was true for the late gestational exposed individuals

and their sibling controls. Data quality control and filtering were done as previously described [60]. Data filtering consisted of the removal of CpG dinucleotides of which the measurement could be confounded by single nucleotide polymorphisms and of CpG dinucleotides of which the measurement success rate was below 80%. Common causes of a lower success rate include fragments bordering on the upper and lower limits of the mass range that can be detected and cases of fragments of which the base of the peak signal in the mass spectrum overlapped another fragment. The success rate for the CpG containing fragments that could be measured within the limits of the methodology was 93.3%.

Statistical analysis

The analyses were performed within sibships to minimize the possible confounding effects of differences in familial environment and genetic. We applied linear mixed models on the raw data without imputation of missing values to calculate exposure specific differences between sibling pairs. All the analyses account for family relations, age at examination, bisulfite plate and the correlation between CpG dinucleotides. Sibship was entered as a random effect. Age at examination, bisulfite plate, exposure status and CpG dinucleotide, and where appropriate sex, were entered as fixed effects. The test for the timing specificity of the association between exposure and DNA methylation levels was done by adding a variable indicating the timing of the exposure and merging the periconceptional and last trimester datasets: timing specificity was tested by adding an interaction term of exposure status times the timing of the exposure to the linear mixed model. Testing the sex specificity of the associations between famine exposure and DNA methylation levels was done by adding a term for sex times the exposure status of the individuals in the linear mixed model.

The linear mixed model was chosen over a standard paired t-test because it allows for the analysis of multiple individual CpG dinucleotides in one test, accounts for the correlation between adjacent CpG dinucleotides, includes relevant adjustments within the model on the raw data, and uses all available data. The linear mixed model reduces to a test with identical outcome to a paired t-test if the within

family difference is assessed for a single CpG dinucleotide, if no adjustments are performed and if there are no incomplete data for the sib pairs.

The basal difference in DNA methylation between men and women in the controls was calculated using ANOVA. All P-values reported are two-sided. All analyses were performed using SPSS 16.0.

Acknowledgements

We thank the participants of the Hunger Winter Families Study, TNO Quality of Life for contact tracing, the staff of the Gerontology and Geriatrics Study Center at the Leiden University Medical Center for performing the clinical examinations and the Central Clinical Chemical Laboratory for extracting genomic DNA. This work was supported by grants from the Netherlands Heart Foundation (2006B083 to B.T.H); the U.S. National Institutes of Health (R01-HL067914 to L.H.L); The Netherlands Organization for Scientific Research NWO (911-03-016 to P.E.S); and the European Union funded Network of Excellence LifeSpan (FP6 036894). Dr. Lumey was supported by a 2008-2009 Lorentz fellowship from the Royal Netherlands Academy of Sciences.

Conflicts of Interest statement. No conflicts of interest declared.

Table S3.1. Details of the measured amplicons and the PCR primers

Gene Name	Genomic location (NCBI b36.1)	Function	Primer forward [b]
(alias)	**CpG sites analyzed [a]**	**Features / (genetic) associations**	**Primer reverse [c]**
IL10	chr1:205012634-205012962	Anti-inflammatory	TGATTGGTTGAATATGAATTTTTGTAT
	CpG 1, 2&3, 4	Promoter, MIR retrotransposon, CpG 4 is an ATF6, XBP1, PPARG binding site / atherosclerosis, schizophrenia	CACCCCCTCATTTTTACTTAAAAA
GNASAS	chr20:56859210-56859503	Imprinting control region GNAS locus	GTAATTTGTGGTATGAGGAAGAGTGA
(*NESPAS*)	CpG 1&2, 3&4, 6, 7, 8&9, 10-12, 13&14, 15, 17-19	promoter / Pseudohypoparathyroidism 1B, QTL blood pressure	TAAATAACCCAACTAAATCCCAACA
INSIGF	chr11:2138912-2139216	Embryonic growth, metabolism	GTTTTGAGGAAGAGGTGTTGA
(*INS*)	CpG 2, 4, 5, 6	Promoter, imprinted / SGA, *IGF2* levels in umbilical cord	ACCTAAAATCCAACCACCCTAA
LEP	chr7:127668290-127668646	Appetite regulation & fat metabolism	GTTTTTGGAGGGATATTAAGGATTT
	CpG 1, 16&17, 19-21, 22, 25, 27	Promoter, CpG 22 is an C/EBP binding site controlling basal gene expression / birth weight, BMI, blood pressure	CTACCAAAAAAAACCAACAAAAAAA
MEG3	chr14:100361166-100361395	Embryonic growth inhibitor	TTTTTTTTAATAGTATTTTGATTTTTG
(*GTL2*)	CpG 2, 3, 4, 5&6, 8&9, 10&11	Promoter, CTCF binding site, part of DLK1-DIO3 imprinting control mechanism	AAATAATCCCCACACACATACC
ABCA1	chr9:106730323-106730642	Cholesterol transport	ATTTTATTGGTGTTTTTGGTTGT

Table S3.1 (continued). Details of the measured amplicons and the PCR primers

Gene Name	Genomic location (NCBI b36.1)	Function	Primer forward [b]
(alias)	**CpG sites analyzed [a]**	**Features / (genetic) associations**	**Primer reverse [c]**
	CpG 1, 3&4, 6-9, 10-13, 15&16, 17&18, 19-21, 22&23, 24, 25	Promoter, CpG 10-13 Arnt & USF1 binding site / atherosclerosis, cholesterol levels	ATCAAAACCTATAC-TCTCCCTCCTC
IGF2R	chr6:160346346-160346595	Embryonic growth, apoptosis	AGGTAGAAAAAG-GTTTTGGAAG
	CpG 4&5, 8-10, 11-13, 20&21	Putatively imprinted human IGF2 DMR2 / birth weight	CAAATCTTAAAAAC-TAACTAAAAACC
APOC1	chr19:50109726-50110115	Lipid metabolism	GGAGGAGGGAGAT-TAATATTAATTTGT
	CpG 1, 2, 3, 4, 10, 11	Intron & exons / coronary artery disease	ACCCCAAACCTATA-ACCACCTT
FTO	chr16:52383225-52383575	Development, nucleic acid demethylation	GTTTGTAATTTTAG-TATTTTGGGAGGT
	CpG 8&9, 10&11, 17, 19	Linkage area GWAS / BMI, Diabetes	TTTATTTCCATTTATC-CATTCTCAAA
KCNQ1OT1	chr11:2677737-2678040	Imprinting control region 11p15.5	TTTGGTAGGATTTT-GTTGAGGAGTTTT
(*KvDMR1*)	CpG 1, 6, 8&9, 10-12, 15, 16, 17&18, 20, 21, 24, 25, 26&27	Minimal repressor, promoter / Diabetes, Beckwith-Wiedemann syndrome, Silver Russell syndrome	CTCACACCCAAC-CAATACCTCATAC
NR3C1	chr5:142763741-142764104	HPA-axis	GATTTGGTTTTTTT-GGGG
(*GR*)	CpG 4, 7&8, 9, 10&11, 14, 15&16, 17-20, 31, 33&34, 42	Exon I7, CpG 7 is part of a NGF1-A binding site / depression	TCCCTTCCCTA-AAACCT
TNF	chr6_qbl_hap2:2790712-2791113	Pro-inflammatory	GGGTATTTTTGAT-GTTT-GTGTGTT

Table S3.1 (continued). Details of the measured amplicons and the PCR primers

Gene Name (alias)	Genomic location (NCBI b36.1) CpG sites analyzed [a]	Function Features / (genetic) associations	Primer forward [b] Primer reverse [c]
	CpG 1-3, 5, 9, 10, 11	Promoter / Graves' disease, asthma	CAATACTCATA-ATATCCTTTC-CAAAAAA
GRB10	chr7:50818080-50818483	IIS inhibitor	GGAATTTTAGGATTAA-ATTTATGTGA
	CpG 7, 8, 17, 22&23, 24, 25	Promoter, CpG island / diabetes type 2, birth size	AACTTCCAAAAA-AAACCTCTCC
CRH	chr8:67253246-67253686	Stress response/ HPA-axis	TGGTTGTT-GTTTTTTTGGTAGG
(*CRP, CRF*)	CpG 1, 2, 9, 10	Promoter, CpG 2 and 10 are TFBS / length of pregnancy, HPA axis	AATTTCTCCACTC-CAAAACCTAAA
GNAS (A/B)	chr20:56896823-56897145	Growth/Lypolytic processes	ATGATTTAATTA-AGGTTTTAGGAAAGG
	CpG 1, 3&4, 7, 8, 9&10, 12, 13-15, 16-19	Promoter / blood pressure QTL	TAAAAATA-CAAAACCT-CCCCTACTC

a: CpG sites measured in each amplicon. Counting starts from the part of the amplicon that has the identical sequence of the forward primer. Multiple CpG sites that are named together between brackets were not individually resolved by fragmentation in the mass spectrometer and thus measured simultaneously.

b: Forward primer that will amplify the bisulfite converted genomic DNA. For the Epityper methodology a 10mer spacer tag is added at the 5' primer end with the following sequence: 5'-AGGAAGAGAG+primer

c: Reverse primer that will amplify the bisulfite converted genomic DNA. For the Epityper methodology a T7 promoter is added to the 5' primer end with the following sequence: 5'-CAGTAATACGACTCACTATAGGGAGAAGGCT+primer

Abbreviations:

MIR: type of retrotransposon
QTL: quantative trait locus
SGA: small for gestational age
BMI: body mass index
TFBS: transcription factor binding sites
HPA: hypothalamus-Pituitary axis
IIS: insulin signaling
NGF1-A: a transcription factor reported to influence *NR3C1* expression [93]
GWAS: genome wide association study

Chapter 4

Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs

Rudolf P Talens,[1] Kaare Christensen,[2,3,4] Hein Putter,[5] Gonneke Willemsen[6], Lene Christiansen,[2,3,4] Dennis Kremer,[1] H. Eka D. Suchiman,[1] P. Eline Slagboom,[1,7] Dorret I. Boomsma,[6] Bastiaan T. Heijmans[1,7]

1. Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
2. The Danish Aging Research Center and The Danish Twin Registry, University of Southern Denmark, Odense C, Denmark
3. Department of Clinical Genetics, Odense University Hospital, Odense C, Denmark
4. Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense C, Denmark
5. Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
6. Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands
7. Netherlands Consortium for Healthy Ageing.

Published in Aging Cell, Volume 11, Issue 4, pages 694 – 703, 2012

Abstract

The accumulation of epigenetic changes was proposed to contribute to the age-related increase in the risk for most common diseases. In this study on 230 monozygotic twin pairs (MZ pairs), aged 18 to 89 years old, we investigated the occurrence of epigenetic changes over the adult lifespan. Using mass spectrometry, we investigated variation in global (LINE1) DNA methylation and in DNA methylation at *INS*, *KCNQ1OT1*, *IGF2*, *GNASAS*, *ABCA1*, *LEP*, and *CRH*, candidate loci for common diseases. Except for *KCNQ1OT1*, inter-individual variation in locus specific DNA methylation was larger in old individuals than in young individuals, ranging from 1.2 fold larger at *ABCA1* ($p = 0.010$) to 1.6 fold larger at *INS* ($p = 3.7 * 10^{-07}$). Similarly, there was more within-MZ-pair discordance in old as compared with young MZ pairs, except for *GNASAS*, ranging from an 8 % increase in discordance each decade at *CRH* ($p = 8.9 * 10^{-06}$) to a 16 % increase each decade at *LEP* ($p = 2.0 * 10^{-08}$). Still, old MZ pairs with strikingly similar DNA methylation were also observed at these loci. After 10 year follow-up in elderly twins, the variation in DNA methylation showed a similar pattern of change as observed cross-sectionally. The age-related increase in methylation variation was generally due to unique environmental factors, except for CRH, for which familial factors may play a more important role. In conclusion, sustained epigenetic differences arise from early adulthood to old age and contribute to an increasing discordance of MZ twins during ageing.

Key words

Epigenetics; Aging; MZ twin design; Full adult lifespan; DNA methylation; Stochastic variation

Introduction

The risk for most common diseases increases with age. A lifetime of accumulated epigenetic changes was proposed to contribute to the development of such diseases [186]. Epigenetic mechanisms determine the expression potential of genes without changing the DNA sequence [29]. The molecular basis includes the methylation of cytosines in CpG dinucleotides, which together with histone modifications, non-coding RNAs, and localization influence the accessibility of a genomic locus to the transcriptional machinery [23,25]. DNA methylation can be measured on DNA samples that are commonly available in biobanks [195].

Various studies have investigated whether DNA methylation can change with increasing calendar age. A cross-sectional study of limited sample size reported the genome-wide absence of changes in mean DNA methylation between young (26 years) and old (68 year) individuals [196]. However, cross-sectional studies that focus on changes in mean DNA methylation can only detect age-related changes that are in the same direction for most individuals. A cross-sectional study that focussed on DNA methylation at the *COX7A1* locus reported greater inter-individual variation in 20 elderly individuals (> 60 years old) compared with 20 young individuals (< 30 years old) [110], indicating that DNA methylation can indeed change with age in a direction that differs per individual.

Longitudinal studies are even better suited to investigate this type of age-related methylation changes, even though most of them rarely span more than a period of 2 decades. A study on global DNA methylation, a measure of the average methylation level of (a representative portion of) CpG sites across the genome, observed changes with a direction that was individual-specific in whole blood samples of 111 individuals (59-86 years) followed over 11 year and 127 individuals (5-72 years) followed over 16 years [131]. Yet, two smaller studies with 10-20 years follow up demonstrated that at specific genomic loci DNA methylation in blood and buccal swab samples can remain remarkably stable [195,197].

A particularly powerful design to investigate the accumulation of epigenetic changes with age are studies of monozygotic twins [198]. MZ co-twins have the same age and a virtually identical genotype, thus controlling for their effect

on DNA methylation [60,92], while they may differ in their lifetime exposure to environmental factors and can develop small phenotypic differences with age [199]. To our knowledge only one study has as yet adopted this design to study age-related changes into adulthood. This study on 40 MZ pairs aged 3 to 74 year reported that older MZ pairs (> 28 years old) showed larger within-pair epigenetic differences than younger MZ pairs (< 28 years old) in total DNA methylation and total histone acetylation levels. Although the functional relevance of such measures is uncertain, analysis on smaller subsets of MZ twins indicated similar trends at sites throughout the genome, most of which were repetitive sequences, but also included singly copy genes [130]. Also, it remains unclear at what age such changes start to arise in the population and at what rate they subsequently progress with increasing age.

The studies performed thus far were generally relatively small, focused on measures of average methylation of the genome and/or could investigate limited periods of the adult lifespan only. Here, we report on age-related changes in locus-specific DNA methylation in a combined cross-sectional and longitudinal study on 460 individuals comprising 230 MZ pairs aged 18 to 89 years. The long age-range investigated allowed us to study whether epigenetic changes accumulate linearly, exponentially or in bursts during the full adult life span. Furthermore, we evaluated the influence of familial versus individual factors on the age-related increase in discordance. We assessed both global DNA methylation with the LINE1 assay [115], and locus-specific DNA methylation close to genes implicated in various age-related diseases, namely the non-imprinted loci *LEP*, *ABCA1*, *CRH*, and the imprinted loci *IGF2*, *INS* (alternate symbol *INSIGF*), *KCNQ1OT1* (alternate symbol *KVDMR*), and *GNASAS* (alternate symbol *NESPAS*). These loci were selected on the basis of their previously shown features of epigenetic regulation as observed in human, animal, or cell culture experiments [133,141–143,145,200,201] (Table 4.1).

Results

DNA methylation in young and old MZ twins

Means and inter-individual variation

Global DNA methylation and methylation status at 9 specific loci was compared between young twins (n = 132 individuals)

Table 4.1: Characteristics of the methylation assays and their corresponding loci

Locus (alias)[a]	Candidate for	Assay on	Literature[b]	CpG island	Imprinted	CpGs[c] units	CpGs[c] sites
ABCA1[d]	Cholesterol transport	Proximal promoter	± [143]	+		10	18
CRH[e]	Stress response	Proximal promoter	+ [142]			7	7
GNASAS (*NESPAS*)[e]	cAMP-dependent pathway	Antisense promoter	+ [201]		+	10	18
IGF2[e]	Growth and insulin signalling	DMR	+ [133]		+	4	5
IGF2-pter	Growth and insulin signalling	pter of DMR	+ [133]		+	5	5
IGF2-qter	Growth and insulin signalling	qter of DMR	+ [133]		+	7	8
INS (*INSIGF*)	Glucose metabolism (placental development)	Proximal promoter	+ [200]		+	5	5
KCNQ1OT1 (*KVDMR*)[e]	Development, growth, and metabolism	Antisense promoter	+ [145]	+	+	11	15
LEP[e]	Metabolism	Proximal promoter	+ [141]	+		7	10
LINE1[e]	Proxy for global DNA methylation	5' promoter (conserved)	± [115]			8	11

a: Loci are given in alphabetical order

b: CpG-site methylation previously reported to associate with gene expression, ± means that this association is hinted at

c: Amount of CpG units and CpG sites as measured in this study

d: For ABCA1 only the methylated CpG sites at the 5' end of the assay are used in analyses [135]

e: Methylation of these assays was investigated over the full adult lifespan and over a 10 year follow-up period in elderly twins

Table 4.2: Average DNA methylation in young and old individuals

	Mean % methylation (SD)		Difference of mean, % methylation		Fold change of variation	
Locus	**Young**	**Old**	**Old – Young (SE)**	**P_{Mean}**	**SD_{Old} / SD_{Young}**	**$P_{variation}$**
Global	61.3 (1.2)	60.9 (1.8)	– 0.4 (0.19)	.033	1.5	$2.3 * 10^{-05}$
KCNQ1OT1	31.2 (2.3)	29.1 (2.3)	– 2.1 (0.28)	$2.4 * 10^{-12}$	1.0	.999
GNASAS	47.3 (3.2)	52.7 (4.4)	+ 5.4 (0.46)	$4.1 * 10^{-25}$	1.3	.002
ABCA1	34.9 (7.8)	38.5 (9.6)	+ 3.6 (1.04)	$6.2 * 10^{-04}$	1.2	.010
INS	81.2 (2.3)	77.6 (3.9)	– 3.6 (0.38)	$9.8 * 10^{-18}$	1.6	$3.7 * 10^{-07}$
IGF2DMR	54.7 (3.8)	51.2 (4.9)	– 3.5 (0.52)	$2.8 * 10^{-10}$	1.3	.002
IGF2_qter	70.9 (2.8)	68.5 (5.4)	– 2.4 (0.52)	$1.3 * 10^{-05}$	1.9	$4.6 * 10^{-07}$
IGF2_pter	45.7 (3.0)	42.1 (4.4)	– 3.6 (0.45)	$6.5 * 10^{-14}$	1.5	$3.5 * 10^{-05}$
LEP	27.6 (4.2)	32.2 (6.4)	+ 4.5 (0.65)	$3.6 * 10^{-11}$	1.5	$1.3 * 10^{-05}$
CRH	63.4 (4.3)	63.7 (6.6)	+ 0.3 (0.67)	.671	1.5	$2.2 * 10^{-06}$

and old twins (n = 134 individuals). Old individuals had slightly lower mean global DNA methylation than young individuals (Table 4.2). Larger differences were observed at specific loci. Old individuals had lower mean DNA methylation at 5 of 9 loci (*INS*, *KCNQ1OT1*, and the 3 adjacent *IGF2* loci; Table 4.2) and showed higher mean DNA methylation at 3 loci (*LEP*, *ABCA1*, and *GNASAS*). No difference was observed for *CRH*.

The inter-individual variation in global DNA methylation expressed as the standard deviation (SD) was small in both age groups (SD_{young} = 1.2 % and SD_{old} = 1.8 %; Table 4.2). At specific loci, it ranged from small in both age groups at *KCNQ1OT1* (SD = 2.3 % in both age groups) to large in both age groups at *ABCA1* (SD_{young} = 7.8 % and SD_{old} = 9.6 %; Table 4.2). With the exception of *KCNQ1OT1*, methylation

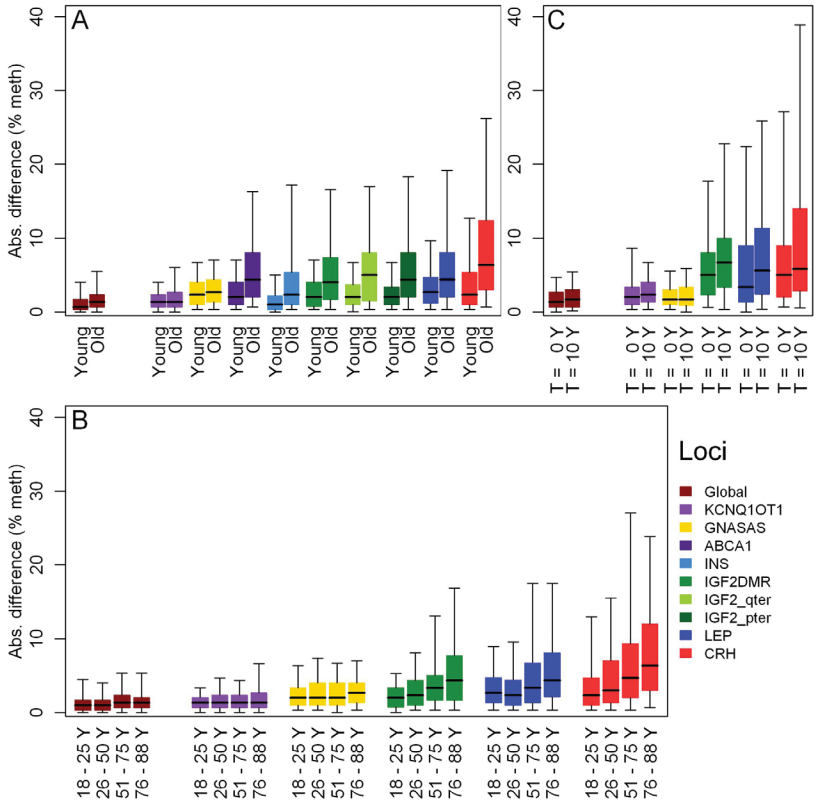


Figure 4.1: Increase in within-pair methylation discordance with age. Absolute within pair difference in % DNA methylation for global DNA methylation and for specific loci plotted for **A.** Young MZ twins (under 30 years old; n = 66 pairs) vs. old MZ twins (over 74 years old; n = 67 pairs). **B.** The full age range divided by the 4 stages of adult life; from left to right: up to 25 years old (early adulthood, n = 30 pairs), 26 up to 50 years old (early middle-age, n = 78 pairs), 51 up to 75 years old (late middle age, n = 56 pairs), and over 76 years old (old age, n = 54 pairs). **C.** In 19 elderly MZ pairs (73 – 82 years old) after 10 years follow-up (83 – 92 years old). The bars show the inter quartile range, the thick line in the center of the bar shows the median and the whiskers show the 5th (bottom whisker) and 95th (top whisker) percentiles.

variation was always larger in old individuals than in young individuals, irrespective of age-related differences in mean DNA methylation. The SD of global DNA methylation was 1.5 fold larger in old individuals ($P = 2.3 * 10^{-05}$). At specific loci the age-related difference ranged from 1.2 fold larger SD at *AB CA1* ($p = 0.010$) to 1.6 fold larger SD at *INS* ($p = 3.7 * 10^{-07}$; Table 4.2).

Within-pair discordance

The extent of within-pair methylation discordance was also compared between the young and old twins. Similar to the inter-individual variation, a small within-pair discordance in global methylation was observed in both age groups. At specific loci it ranged from small in both age groups at *KCNQ1OT1* to large for both age groups at *CRH*. Furthermore, the absolute within-pair discordance in old MZ pairs was always greater than in young MZ pairs (Figure 4.1A). Notwithstanding this overall increase, the old age group still contained pairs who had strikingly similar DNA methylation. With the exception of *GNASAS*, the SD of the within-pair differences, quantifying group discordance, was significantly higher in old as compared with young MZ pairs (Table 4.3). For global DNA methylation, methylation discordance in the old MZ pairs was almost double that of the young MZ pairs ($p = 9.8 * 10\text{-}05$). At specific loci the increase in discordance ranged from 1.4 fold greater in old MZ pairs at KCNQ1OT1 ($p = .005$) to 2.7 fold greater at ABCA1 ($p = 3.8 * 10\text{-}07$).

DNA methylation across the complete adult lifespan

Changes in within-pair methylation discordance between different age categories

The timing of the occurrence of age-related changes in methylation discordance during adult life was investigated in 219 MZ pairs aged 18 to 89 years old, including an additional 61 middle aged MZ pairs (30 – 65 years old) and 25 old MZ pairs (> 65 years old; Table 4.5). In this extended set of MZ twins, global DNA methylation and methylation at 5 specific loci, representative of the 9 loci studied in the young and old MZ pairs as described before, were measured (Table 4.1). The observed absolute discordance was plotted for the four stages of adult life: early adulthood (up to 25 years old, n = 30 pairs); young to middle-age (26 years up to 50 years old, n = 78 pairs); middle-age to seniority (51 years up to 75 years old, n = 56 pairs); and old age (over 76 years old, n =

Table 4.3: Variation of within-MZ-pair methylation difference in young and old MZ pairs

	Estimate of variation as SD (% methylation)[a]		Fold change of variation	
Locus	Young MZ pairs	Old MZ pairs	Old / young	$P_{variation}$
Global	1.1	2.1	1.9	$9.8 * 10^{-05}$
KCNQ1OT1	1.9	2.6	1.4	.005
GNASAS	3.1	3.4	1.1	.193
ABCA1	2.8	7.7	2.7	$3.8 * 10^{-07}$
INS	2.1	4.6	2.2	$2.3 * 10^{-06}$
IGF2DMR	2.6	4.9	1.9	$2.1 * 10^{-05}$
IGF2_qter	2.4	5.9	2.4	$9.7 * 10^{-07}$
IGF2_pter	3.3	5.9	1.8	.002
LEP	3.5	7.3	2.1	$1.0 * 10^{-05}$
CRH	4.6	7.7	1.7	$2.3 * 10^{-04}$

a: Mean within-MZ-pair difference at any locus did not significantly deviate from 0 at in both age groups

54 pairs). Discordance in global DNA methylation was small in all age groups (Figure 4.1B). At the loci *IGF2*, *LEP*, and *CRH* an increase in methylation discordance across the age-groups was observed, starting from young adulthood onward at *IGF2* and *CRH*, and from middle age onward at *LEP*. Age related changes in discordance were less apparent at *GNASAS*, and *KCNQ1OT1* (Figure 4.1B). Plotting the absolute within-pair methylation differences at individual CpG units against age yielded similar results (Supplementary figures S4.1 to S4.6).

To quantify these observed changes the proportional increase in discordance was also estimated per decade. For global DNA methylation within-pair discordance was 9 % greater each decade ($p = 3.4 * 10^{-06}$, Table 4.4). The greatest increase in within-pair discordance was found at *LEP* with 16 % greater discordance each decade ($p = 2.0 * 10^{-08}$), at *IGF2* and *CRH* within-pair discordance increased each decade with 11 % ($p = 3.8 * 10^{-09}$), and 8 % ($p = 8.9 * 10^{-06}$) respectively. At *GNASAS* discordance did not change significantly with age. At *KCNQ1OT1*, although the discordance was small at all ages, it did increase by 11 % each decade (p = .002; Table 4.4). These observed relative increases in discordance were confirmed with absolute differences in the amount of

Table 4.4: Variation of within-pair methylation discordance over the full adult lifespan

Locus[a]	Baseline estimate of variation as SD (% methylation)[b]	Increase in variation per decade	
		Proportional increase[c]	P[d]
Global	2.2	9.1 %	$3.4 * 10^{-06}$
KCNQ1OT1	1.9	11.4 %	.002
GNASAS	3.5	2.7 %	.415
IGF2DMR	4.4	11.1 %	$3.8 * 10^{-09}$
LEP	4.8	16.0 %	$2.0 * 10^{-08}$
CRH	8.4	8.0 %	$8.9 * 10^{-06}$

a: Loci are ordered from top to bottom as in Figure 1B from left to right

b: The SD of discordance at baseline, estimated from the residual variance and the random effect intercept of the linear mixed model

c: The proportional (in percentage) increase in variation of discordance each decade, estimated from the random effect of age

d: One-sided p-value from a Z-test on the random effect estimate of age from the same linear mixed model

discordance when testing the homogeneity of variance in discordance across the age groups at individual CpG units (Supplementary table S4.7).

Dutch and Danish MZ pairs were investigated in this study. The observation that the increase in within-pair twin discordance was not exclusive to the old age group (predominantly Danish twins) but was already apparent at younger age (Dutch twins) indicated that geographical origin of the MZ pairs did not contribute to our findings. To further exclude the influence of origin, we tested whether age-related changes in methylation variation or discordance were different between old Dutch (n = 25 pairs) and Danish (n = 67 pairs) twins. No significant influence of country on age-related changes was observed for either variation or discordance (Supplementary table S4.2).

The influence of cellular heterogeneity on methylation variation, discordance and their age related increases

Methylation was measured on genomic DNA extracted from whole blood and variation in cellular heterogeneity could induce differences in DNA methylation. In the young twins, neither variation in global DNA methylation (n = 132 individuals) nor within-pair differences in global methylation (n = 66 pairs) were associated with cellular heterogeneity,

approximated by percentage neutrophils, as recently described [195]. The same was true for the majority of loci (at 6 out of 9 loci for DNA methylation and at 5 out of 9 loci for within-pair differences; Supplementary table S4.3). The strongest influence of cellular heterogeneity was observed for *LEP* ($p = 1.2 *10^{-10}$ and $p = 1.0 *10^{-06}$, for both tests respectively). Even in this case, 85 % of inter-individual variation in *LEP* methylation and 80 % of within-pair differences were independent of cellular heterogeneity.

The composition of the leukocyte population changes with age, of which the contribution to our findings was investigated in all Dutch twins. Age-related changes in inter-individual variation (n = 304 individuals) at 3 out of 5 loci, and in within-pair discordance (n = 152 pairs) at 4 out of 5 loci were not associated with changes in cellular heterogeneity (Supplementary table S4.4). Although an association with cellular heterogeneity was observed ($p = .006$ and $p = 8.2 * 10^{-04}$, for both tests respectively), 90 % of the age-related changes in either global methylation variation or global methylation discordance was not attributable to it. The strongest influence was observed at *LEP* ($p = 4.1 * 10^{-18}$ and $p = 4.2 * 10^{-12}$, for both tests respectively), yet 90 % of change in inter-individual variation, and 85 % of change in within-pair discordance were not attributable to cellular heterogeneity. Thus, the age-related changes in DNA methylation observed cannot be explained by changes in leukocyte population composition.

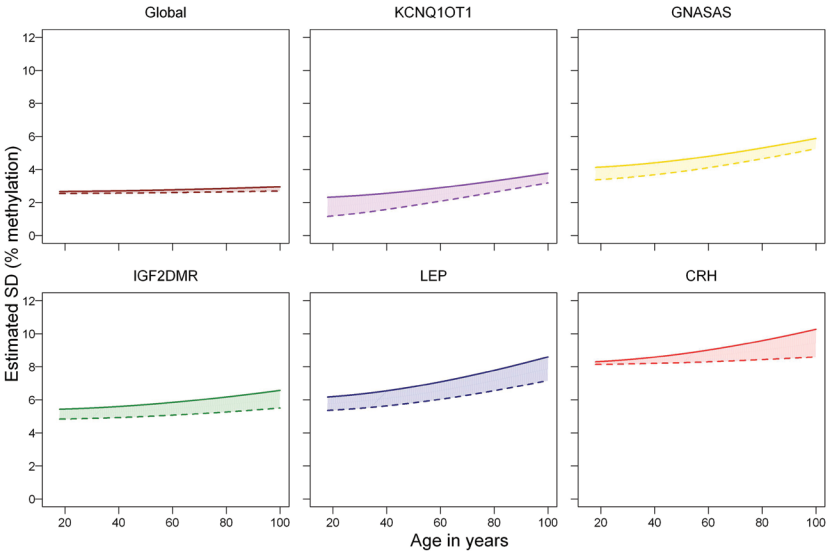


Figure 4.2: Contribution of familial and unique (individual) environmental factors to increasing methylation variation with age.
Estimated changes in the variation in DNA methylation (y-axis, given as standard deviation in percentage DNA methylation) plotted against the adult lifespan (x-axis, age in years). The changes in total variation of DNA methylation (area under the thick line), and the contributions of familial factors (shared environment and genotype; filled in area between thick and dashed line) and unique factors (individual environment; blank area under the dashed line), were estimated in 219 MZ pairs ranging from 18 to 89 years old. From top left to bottom right, global DNA methylation and the loci *KCNQ1OT1*, *GNASAS*, *IGF2DMR*, *LEP*, and *CRH* are investigated. Significance of the estimates of total, familial, and individual related increases in variation are given in Supplementary table S4.6.

Longitudinal changes in DNA methylation in old age

Inter-individual and within-pair epigenetic variation, global and at the same 5 loci, were also investigated in 19 elderly twin pairs during 10 years follow-up (DNA samples obtained in 1997 and 2007). Global and locus specific inter-individual methylation variation was modestly larger after 10 year follow-up, except at *CRH* (Supplementary table S4.5). Changes in within-pair methylation discordance showed a similar pattern as observed in comparing changes from young to old adults (Figure 4.1: compare C with A and B). Global discordance and that at 3 loci (*IGF2DMR*, *LEP*, and *CRH*) had increased during the follow-up period, whereas no

change was observed at the remaining 2 loci (*KCNQ1OT1* and *GNASAS*).

Familial and unique environmental factors

The study of MZ twin pairs enables separation of the effects of familial (i.e. genetic and common environment) and unique environmental environment on the accumulation of DNA methylation differences with age. For global methylation, the total variation was relatively small and mainly due to unique environment. The increase in variation with age was limited but significant ($p = .004$) and mostly attributable to unique environment ($p = .003$; Figure 4.2, top left graph; Supplementary table S4.6). The total variation increased significantly with age at all loci in line with the previous analyses (Supplementary table S4.6). This increase could be mainly attributed to unique environmental factors except for the age-related increase in variation at CRH methylation which had a familial component ($p= 0.007$).

Discussion

In this study we report sustained age-related increases in variation of DNA methylation, in an analysis of 460 individuals, comprising 230 MZ pairs, aged from 18 to 89 years. Previously this question was investigated using crosssectional and longitudinal study designs on smaller sample sizes with narrower age ranges [108,130,195,197]. Our study extends their findings over the full adult lifespan, supporting the notion that a gradual accumulation of epigenetic changes, globally and at imprinted and non-imprinted loci, occurs up to very old ages. How such changes affect gene expression remains unclear, although some evidence suggests that small differences in DNA methylation may cause an amplified effect on gene expression [188]. The increase in epigenetic variation was mainly due to unique individual factors that cover both stochastic processes and environmental exposures, between which the design of our study cannot distinguish. This may lead to age-related epigenetic dysregulation and may contribute to the age dependency of common diseases [29,186]. However, studies that can appropriately address the latter hypothesis will be complex in their design and execution due to the relatively small effect sizes involved and the tissue and cell specific

nature of age-related changes [202].

We measured LINE1 methylation to assess global DNA methylation [115] and found that the variation in global DNA methylation, both inter-individual and within-pair, was small at all ages, but increased proportionally with age, in accordance with a longitudinal study over ten years [131]. The small amount of global variation observed may relate to the fact that global DNA methylation assays measure methylation at a multitude of similar loci distributed throughout the genome [115,163], while the introduction of stochastic changes occurs at individual loci. We observed the most prominent age-related changes in variation at specific candidate loci for common age related diseases, such as *IGF2*, *LEP*, and *CRH*, of which the expression was shown to be influenced by its DNA methylation [133,141,142]. We observed a substantial difference between the loci in their vulnerability to epigenetic drift, which was consistently found in our various analyses (inter-individual variation, within-pair discordance, young versus old individuals, the adult lifespan, or a longitudinal analysis in old subjects). Of interest, the greatest age-related changes were observed for non-imprinted loci, whereas imprinted loci appeared more stable.

MZ twin pairs share characteristics such as age, sex, genotype and their developmental and childhood environment (e.g. upbringing / education). They may acquire more unique characteristics as they grow older, since different choices on for instance lifestyle and occupation can increasingly change their living environments. The shared characteristics, namely their genotype and shared environment, and their individual characteristics, commonly named the unique environment, may both contribute to epigenetic variation [103,198]. We used the MZ twin study design to investigate how much of the age-related increase in variation is contributed by the familial factors and by the unique individual environment [198,203]. We observed that most of the variation in young adults could be attributed to the individual environment, indicating that DNA methylation is at least partly independent of familial (e.a. genetic) factors. Interestingly, studies on neonate and infant MZ pairs indicated that such differences are already present and indeed increase at a very young age [62,204]. At most loci we found that the age-related changes in variation were also mostly attributable to the unique individual

environment, supporting the idea that age-related changes in DNA methylation may be mostly independent from familial factors [29,59,186]. However, at *CRH* the increase in variation was mostly attributed to familial factors. Residual batch effects seem an unlikely explanation for this observation considering that the design of this study involved a batch allocation scheme that took age and sex into account and the statistical models contained a factorial variable to adjust for its potential influence. A previous study reported familial clustering of variation over time in global DNA methylation using the LUMA assay [131]. Since we studied MZ twins only, we cannot further distinguish whether this familial component is related to genetic or shared environmental factors, both of which may influence the susceptibility to stochastic or environmentally driven changes in DNA methylation [139].

In this study we observed that differences in mean methylation between the young and the old age groups were relatively small or even absent, while increases in variation were generally more substantial. This finding indicates that epigenetic changes accumulating with age are generally non-directional or are the outcome of many smaller directed changes with, in part, opposite direction. Changes in DNA methylation can be stochastic or environmentally driven; the relative contribution of each source of variation cannot be investigated in our study. Stochastic epigenetic changes can occur without any environmental influence and may be related to imperfect DNA methylation maintenance mechanisms [70]. We observed that the age-related increase in variation of DNA methylation was gradual from early adulthood to old age, which is compatible with stochastic effects [186]. Environmentally driven epigenetic changes may occur as a consequence of environmental exposures related to for instance lifestyle and occupation [71,73,108]. It is even conceivable that stochastic epigenetic changes occur more often under certain environmental conditions, parallel to stochastic genetic mutations that occur after exposure to high UV irradiation.

The power of cross-sectional studies in MZ twins is that large age-ranges can be studied in contrast to what is practically possible in longitudinal studies on unrelated individuals. We investigated changes in epigenetic variation, both in terms of inter-individual variation and within-pair difference, over a large age-range. We observed

similar increases in either measure of epigenetic variation throughout the adult life-span. These results are in line with longitudinal analyses over relatively short time spans in old age, as we report here, middle age [195,197], and childhood [204]. In view of the consistency of these findings, it is unlikely that generation effects significantly contributed to our findings. Importantly, we were also able to exclude age-related changes in cellular heterogeneity of whole blood as an explanation for our observations, which was previously proposed as a major concern for the interpretation of such studies [103]. Whether increased methylation variation in blood has any biological (phenotypic) relevance is an important and as yet unresolved question, although it is likely that other tissues than blood are similarly affected by epigenetic drift [108,110,204]. In general, tissues with a high rate of cell division may display more age-related epigenetic variation through stochastic errors in maintaining and transmitting epigenetic information than tissues with lower rates cell division [205].

In this study we investigated DNA methylation of MZ twin pairs with ages distributed across the full adult lifespan. In the first phase of the study, we explored differences by comparing young adult twins (18-30 years) with old twins (> 74 years) to generate sufficient contrast between groups. Such age-groups at either extreme of the adult lifespan could not be selected from a single twin register. This required a careful consideration of potential biases created by selecting young twins from the Netherlands and old twins from Denmark and additional measurements to confirm the validity of our findings. Genetic differences between the populations were unlikely to play a role, since a genome-wide analysis of SNPs of Northern European countries reported a similar structure of genetic variation in the two countries [206]. Further, similar procedures were used for drawing blood [116,207], DNA was extracted using standard protocols, and there was no indication for differences in DNA quality (including OD 260/280 measurements, bisulfite conversion rate and success rate of DNA methylation assays). Moreover, if DNA quality was different between the populations, one would expect a similar systematic effect on all assays, which could be taken into account in our statistical analysis. In contrast, the loci we studied showed a substantial variation in the degree to which the DNA methylation was higher in old twins and was absent

for *KCNQ1OT1*. More importantly, we experimentally validated the finding from the first phase for 6 loci. In this second phase, we compared the old Danish twins with a subset of Dutch twins specifically selected for a maximum age overlap (within the limitations of the availability of old twins in the Netherlands twin register). We found no indication for DNA methylation differences between Dutch and Danish twins with a similar old age. Furthermore, the locus-specific associations of DNA methylation variation with age originally observed in the young-old comparison were confirmed in both an investigation of intermediate age ranges selected from the Netherlands twin register and in a longitudinal investigation of old Danish twins. Taken together, differences in geographical origin or technical variability between populations are unlikely explanations for our observations in phase one of our study and the second phase yielded further evidence for the occurrence of sustained epigenetic changes during the adult lifespan.

In this study we demonstrate that epigenetic variation in the population, used as a proxy for stochastic and environmentally driven epigenetic changes in individuals, increases gradually with age up to old age. The rate at which changes are introduced differs between loci and can be considerable at loci regulating transcription of nearby genes. The observed increase was mostly driven by the unique environment. Our results have practical implications for study design in epigenetic studies investigating populations with a large age distribution or a long follow-up time [98,195]. Future research should aim to investigate the relative contribution of stochastic and environmental factors to age-related epigenetic changes and the consequences of these changes for development of common age-related diseases.

Methods and materials

Study population

The samples in this study are taken from Longitudinal Study of Aging Danish Twins (LSADT) of the Danish Twin Registry (DTR) [117] and from the Biobank project of the Netherlands Twin Register (NTR) [116]. In the LSADT study, DNA was extracted using the salting out method, and in the BIOBANK project QIAamp DNA Blood Maxi (QIAGEN, Düsseldorf, Germany) was used. DNA from both sources was of high quality ($260/280_{BIOBANK} = 1.80$; $260/280_{LSADT} = 1.90$)

Table 4.5: Basic characteristics of the different MZ twin pair populations investigated in this study

Population[a]	N (pairs)	No. male pairs	Age in years, mean (range)	No. Assays studied	Cell counts[b]
Young Dutch	66	34	25.2 (18.0 – 29.8)	10	+
Middle-aged Dutch	61	15	46.3 (30.0 – 64.0)	6	+
Old Dutch	25	8	70.5 (65.0 – 78.0)	6	+
Old Danish	67	34	79.3 (74.1 – 89.0)	10	na
Baseline Danish	19	8	76.6 (73.2 – 81.8)	6	na
10 year follow-up	19	8	86.5 (83.4 – 91.8)	6	na
Full adult life span[c]	219	91	52.7 (18.0 – 89.0)	6	

a: The designation for the population sample in this study, Dutch twins are participants in the Netherlands Twin Register (NTR), Danish twins are participants in the Danish Twin Registry (DTR)

b: Availability of data on the amount of the major leukocyte fractions (neutrophils, lymphocytes, monocytes, basophils and eosinophils)

c: The population sample covering the full adult lifespan combines the young, middle-aged and old NTR and the old DTR twins

Selection from the Danish Twin Registry (DTR)

The LSADT study, based on the DTR, is a cohort sequential study of elderly Danish twins. LSADT began in 1995 with an assessment of all members of like-sex twin pairs born in Denmark before 1920. The surviving members were followed up every 2 years and additional cohorts were added at the 1997, 1999, and 2001 assessments and subsequently followed at 2-year intervals. During a home visit in 1997, blood was drawn from 689 individuals, from which DNA was isolated [208]. The LSADT project has been approved by The Danish National Committee on Biomedical Research Ethics (journal VF 20040241). Details on design and data collection were described previously [207].

This study focuses on the monozygotic twin pairs (MZ pairs) of whom DNA was available from the 1997 assessment, (73 years or older; n = 108 pairs). To investigate differences in epigenetic variation between young and old MZ pairs, all 36 male MZ pairs and 37 randomly selected female MZ pairs formed a study population named “old Danish twins”

(Table 4.5). For 2 male MZ pairs and 4 female MZ pairs there was insufficient DNA of both co-twins. These MZ pairs were excluded, the remaining 67 MZ pairs were investigated. For 19 of the LSADT MZ pairs (8 male pairs) a second DNA sample was available from a 10 year follow up in 2007 for both co-twins. Longitudinal epigenetic changes in the elderly were investigated in these MZ pairs, who were named "follow-up Danish twins" (Table 4.5).

Selection from the Netherlands Twin Register (NTR)

In 2004 the NTR started a large scale biological sample collection in twin families to create a resource for genetic studies on health, lifestyle and personality. Between January 2004 and July 2008, adult participants of the NTR (18 years and over) were invited into the project. During a home visit, fasting blood was drawn, from which DNA was extracted and a hematological profile was obtained, consisting of percentages and numbers of neutrophils, lymphocytes, monocytes, eosinophils, and basophils. The study protocol was approved by Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute codes, NTR 03-180). Details on design, biological sampling and data collection were described previously [116].

To investigate differences in epigenetic variation between young and old MZ pairs, 37 MZ pairs of each sex were randomly selected from all NTR MZ pairs who were under 30 years old at sampling (n = 98 pairs, 44 male pairs). For 3 male and 5 female MZ pairs there was insufficient DNA of both co-twins. These MZ pairs were excluded, the remaining 66 MZ pairs were named "young Dutch twins" (Table 4.5). To investigate differences in epigenetic discordance over the full adult life span 37 MZ pairs were selected from all NTR MZ pairs between 30 to 50 years old (135 pairs, 34 male pairs) using a block random selection procedure to guarantee an even distribution over the age range. They were combined with all NTR MZ pairs who were over 50 years old (49 pairs, 16 male pairs). The MZ pairs between 30 and 65 years old were named "middle-aged Dutch twins" (n = 61; 15 male pairs) and the MZ pairs over 65 years old were named "old Dutch twins" (n = 25; 8 male pairs; Table 4.5).

DNA methylation

Assays and measurement

DNA methylation was measured using a quantitatively accurate mass spectrometry-based method (Epityper version 1.05, Sequenom, San Diego, CA, USA) [151,152]. A total of 10 DNA methylation assays were measured in this study, the LINE1 assay for global DNA methylation [115], and 9 assays for DNA methylation at 7 specific genomic loci (*IGF2* (3 assays), *LEP*, *CRH*, *ABCA1*, *INS*, *KCNQ1OT1*, and *GNASAS*; Table 4.1). Two novel assays were designed to assess methylation of the CpG sites directly telomeric, named *IGF2_pter*, and centromeric, named *IGF2_qter*, of the assay at the *IGF2* locus' DMR [60], named *IGF2DMR* for clarity. The primers of each assay were designed to create a PCR bias for completely BS converted DNA [149]. More details on the design, features and measurement of the other 8 methylation assays were described in detail previously [60,115,195]. Briefly: bisulfite (BS) conversion of 0.5 µg of genomic DNA using the EZ 96-DNA methylation kit (Zymo Research, Orange, CA, USA) was followed by PCR amplification (primers are given in Supplementary table S4.1A), fragmentation after reverse transcription and analysis on a mass spectrometer. Fragments that contain one or more CpG sites are called CpG units.

Randomization and quality control

All methylation assays were measured in triplicate on the same bisulphite converted DNA sample. DNA samples of both co-twins of a MZ pair were always allocated to the same batch for BS conversion (on 96-wells plate) and PCR amplification (384-wells plate, 3 x 124 DNA samples). Each batch contained equal proportions of the age groups measured, of the sexes and each PCR batch contained equal proportions of the BS conversion batches. There were two phases of methylation measurement in this study. First, all ten methylation assays were measured in the young Dutch and the old Danish twins, who were randomly divided over the measurement batches. The ten assays contained a total of 74 measurable CpG units, over which 102 CpG sites were distributed (Table 4.1). After quality control [195], 65 CpG units, containing 93 CpG sites remained, with a mean call rate of 96.5 %.

In the second phase, six methylation assays representing observations on the ten assays were measured in the middle-

aged and old Dutch twins, who were randomly divided over the measurement batches, and in the follow-up Danish twins whom were all allocated to a single measurement batch. The six loci contained a total of 47 CpG units, over which 66 CpG sites were distributed. After quality control, 42 CpG units, containing 61 CpG sites remained both for the Dutch twins (average call rate = 96.4 %) and for the follow-up Danish twins (average call rate = 95.7 %). These CpG units were the same that passed quality control in the first phase. Supplementary table S4.1B gives the CpG units and CpG sites that passed quality control of each assay and the call rates per assay in both phases.

Bisulfite (BS) conversion was assessed using the MassArray R package [153], which identifies CpG less fragments containing a TpG and a cytosine on the assay's original genomic sequence. It analyzes the mass spectra treating these fragments as hypothetical CpG sites, since incomplete BS conversion would result in the same mass shift as Cytosine methylation at a CpG site. For both Danish and Dutch twins, this analysis qualified BS conversion as complete within the technical limitation of the method.

Statistical analysis

Definitions

Methylation variation: the Standard deviation of the mean (SD) of the inter-individual differences within a group.

Within-pair methylation difference: the within pair DNA methylation difference at each CpG unit, with DNA methylation of co-twin 1 as the reference: difference = Twin1 – Twin2.

Methylation discordance: the range of the within-pair differences in a group. To quantify age-related changes in discordance, the SD of the within pair differences in a group is used. In figures the absolute within pair differences (absolute discordance) are used.

Days: a continuous variable for the time between the drawing of blood from each co-twin of a twin pair computed in days, with co-twin1 as the reference.

Batch: a categorical variable with a distinct designation for each combination of PCR- and bisulphite batch.

Linear mixed models, description of basic models

Linear mixed models were used to test for age-related changes in DNA methylation of the assays, its variation, and its discordance, as previously described [135,195]. More details on the linear mixed model are given in the supplementary material. In all the linear mixed models used for testing age-related changes in inter-individual methylation variation, DNA methylation was entered as dependent variable. Individual was the subject variable. Necessary adjustments were made by entering age, sex, twin designation (T1 or T2, to account for non-independence), batch, and CpG unit, as fixed effects. In all the linear mixed models used for testing age-related changes in within-pair methylation discordance, the within-pair difference was entered as dependent variable. Family was the subject variable. Necessary adjustments were made by entering age $_{Twin1}$, days, sex, batch, and CpG unit as fixed effects. Both basic models were adapted to suit each specific tests as described below.

Adaptation of basic models for each specific test

DNA methylation, methylation variation, and methylation discordance were compared between young Dutch (n = 66 pairs) and old Danish twins (n = 67 pairs). Age group (young or old) was added to the models as random effect to test for differences in variation or discordance and as extra fixed effect, replacing age, to estimate adjusted group mean methylation or mean within-pair difference and its standard deviation (SD; using the SE of the mean) and test for group differences.

Changes in methylation discordance over the full adult lifespan were tested in all Dutch and the old Danish twins (n = 219 pairs), age was entered as random effect.

Longitudinal methylation variation was investigated in the follow-up Danish twins (n = 19 pairs), the model was adapted as follows: DNA sample (e.a. individual per year of sampling) was the subject variable. Year of sampling (1997 or 2007) was entered as extra random effect and as extra fixed effect. Adjustment for age was done using age at first sampling, no adjustment for batch was required.

Adjusted mean DNA methylation, the differences in means, inter-individual variation and within-pair discordance are all expressed as percentage DNA methylation. The fold

change of methylation variation and discordance between groups is expressed as a proportion by dividing the SD in the older group with the SD in the younger group (SD_{older} / $SD_{younger}$). The change in discordance over the adult lifespan is expressed as the proportional increase each decade as percentage of the discordance of the previous decade. Significance of the age-related changes of variation or discordance was tested with a one-sided Z-test applied on the random effect estimate of age-group, age or sampling time divided by its standard error (SE), which adds up to a Wald test.

Adaptation of basic models for subsidiary tests

To test whether age effects were similar between Dutch and Danish individuals, methylation variation and discordance were compared between old Dutch twins (n = 25 pairs) and old Danish twins (n=67 pairs). Age was entered as extra random effect. An interaction term age*country was entered as extra fixed effect, insignificance of which would establish that Dutch and Danes represent the same population.

Nested linear mixed models were used to investigate confounding by leukocyte population heterogeneity, approximated by percentage neutrophils, as recently described [195]. Confounding of methylation variation and discordance was tested on the young Dutch twins (n = 66 pairs). Confounding of age-related changes was tested on all Dutch twins (n = 152 pairs). The basic models are as described above, with age also entered as a random effect when testing age-related changes. Nested models had percentage neutrophils for testing its influence on methylation variation and the within-pair difference in percentage neutrophils (Twin 1 – Twin 2) for testing its influence on methylation discordance, added to their corresponding basic model as an extra fixed effect. The amount of confounding is determined by the change in residual variance, or the change in the random effect estimate of age, in the nested model with respect to the basic, as described previously [195].

Variance component models for twin analysis

In this study inter-individual methylation variation and within-pair methylation discordance have been investigated separately.

In the statistical models commonly used in twin research both aspects of variation can also be investigated simultaneously, thus correcting each component for the other. The classical twin model for MZ twins [203] postulates that the methylation values (y) at a given locus for co-twins 1 and 2 of twin-pair i are defined by an overall mean (μ) that may depend on age, sex, batch, CpG unit, by a random twin-pair effect (b_i), the familial environment, which stands for the shared factors of the twin pair, including common environment and genotype, and by a residual error (e_{ij}), the individual environment which stands for the factors that are unique to each co-twin (in formula: $y_{ij} = \mu + b_i + e_{ij}$). However, this classical twin model is not able to capture that the methylation variance increases with age. We therefore used the following extension of the classical twin model to allow for such age variation: $y_{ij} = \mu + b_i + e_{ij} + age_{ij} * a_i + age_{ij} * c_{ij}$, with μ, b_i, and e_{ij} as before, and a_i and c_{ij} quantifying shared (familial) and unique (individual) age effects independently from each other and from b_i and e_{ij}. More details on these MZ-twin variance component models [203] are given in the supplementary material.

In the linear mixed models used to test the individual and familial components of variation in DNA methylation over the adult age range (n = 219 pairs), DNA methylation was entered as dependent variable. Individual and family were both subject variables. Age was entered as random effect around family (with the intercept) and as random effect around individual, thereby adjusting each variance component for the other. For necessary adjustments age, sex, batch, and CpG unit were entered as fixed effects. Significance of the age-related increases in total, familial and individual variation were tested with a one-sided Z-test applied on the random effect estimates of age.

The square root of the resulting random effect estimates represents an estimation of the SD, which, expressed as percentage DNA methylation, is easier to interpret. The total of all variation (residual variance, intercept, familial and individual age-related estimate) and the individual variation (residual variance and individual age-related estimate), were plotted against age to visualize the familial and individual age-related increase in variation, since total variance minus individual variance represents the familial variance.

Acknowledgement

The research leading to these results has received funding from the Netherlands Heart Foundation (2006B083), the European Union-

funded Network of Excellence LifeSpan (FP6 036894), the European Union's Seventh Framework Programme IDEAL (FP7/2007-2011) under grant agreement no: 259679, the Netherlands Consortium for Healthy Ageing (NCHA, Grant 05060810) in the framework of the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO), the US National Institute on Aging (P01-AG08761), the VELUX Foundation, NWO (MagW/ZonMW, grants 480-04-004; and 40-0056-98-9032), Spinozapremie (SPI 56-464-14192), CMSB: Center for Medical Systems Biology (NWO Genomics), BBMRI –NL: Biobanking and Biomolecular Resources Research Infrastructure (184.021.007), the VU University: Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam, the European Science Foundation (EU/QLRT-2001-01254), the European Science Council (ERC230374), the National Institute of Health (NIH R01D0042157-01A).

Author contributions

R.P.T., K.C., D.B., P.E.S., and B.T.H. designed the research; K.C., and L.C. designed the LSADT study; G.W., and D.B. designed the NTR biobank project; R.P.T., D.K., and H.E.D.S. performed the research; R.P.T., and H.P. analyzed data; R.P.T., P.E.S., and B.T.H. wrote the manuscript; all authors contributed to interpretation of the data, and critical revision of the manuscript.

Supplementary information on statistical models

Linear mixed models

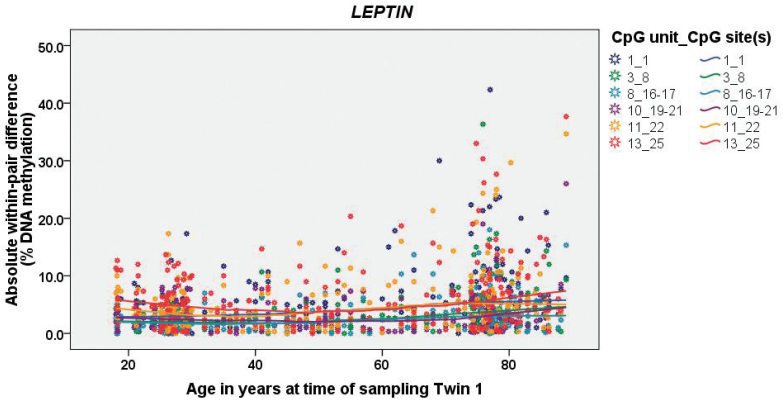
The linear mixed model uses all available methylation data per locus (assay), i.e. methylation of multiple CpG units per locus, accounts for the correlation between methylation of CpG units within a locus, and using this correlation handles data missing at random. It further enables the inclusion of relevant adjustments on the raw data within the same model. For categorical covariates entered as fixed effect the model estimates adjusted means and tests for differences between groups. This reduces to a t-test, when a single CpG unit with complete methylation data is tested without adjustments. For continuous covariates entered as fixed effect the model estimates and tests the adjusted linear relation between the dependent and covariate. This reduces to regression analysis when a single CpG unit with complete data is tested.

For covariates entered as random effect the model estimates the increase in variance of the dependent variable per unit change of the random effect covariate. A Z-test applied on the estimate of variance divided by its standard error (SE), tests the significance of this change in variance of the dependent variable, which in essence adds up to a Wald test.

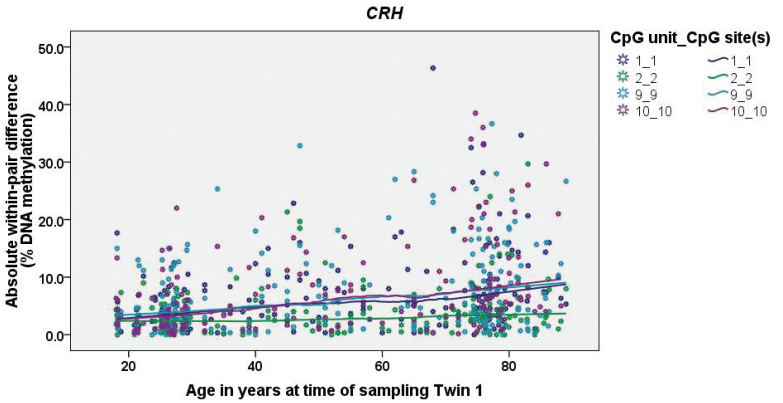
Variance component analysis in twin studies

Define y_{i1} and y_{i1} to be the methylation values at a given locus for twins 1 and 2 of twin-pair $\boldsymbol{i}$. The classical twin model for MZ twins postulates that $\boldsymbol{y_{ij} = \mu + b_i + e_{ij}}$, where μ is an overall mean that may depend on age, sex, batch, CpG unit, $\boldsymbol{b_i}$ is a random twin-pair effect, assumed normal with mean zero and variance σ^2_b, and $\boldsymbol{e_{ij}}$ is the residual error [unique environment], assumed normal with mean zero and variance σ^2_e, and independent of $\boldsymbol{b_i}$. The $\boldsymbol{b_i}$ stands for shared familial factors, including common environment and genotype, and the $\boldsymbol{e_{ij}}$ for unique environment. This model implies that the methylation variance equals $\sigma^2_{tot} = \sigma^2_b + \sigma^2_e$, the correlation between the two methylation values within a twin-pair is the intraclass correlation, given by $\rho = \sigma^2_b / \sigma^2_{tot}$. The within-pair difference $\boldsymbol{y_{i1} - y_{i2}}$ is seen to have a normal distribution with mean zero and variance $\boldsymbol{2\sigma^2_e}$. However, this classical twin model is not able to capture 1) increases of the methylation variance with age, and 2) increases of the variance of the within-pair differences with age. We therefore propose the following extension of the classical twin model, similar to the variance components models in twin analysis [203], to allow for such age related variation: $\boldsymbol{y_{ij} = \mu + b_i + e_{ij} + age_{ij} * a_i + age_{ij} * c_{ij}}$, with μ, $\boldsymbol{b_i}$ and $\boldsymbol{e_{ij}}$ as before, and $\boldsymbol{a_i}$ and $\boldsymbol{c_{ij}}$ mean zero normal random variables with variances σ^2_a and σ^2_c, independent from each other and from $\boldsymbol{b_i}$ and $\boldsymbol{e_{ij}}$. The $\boldsymbol{a_i}$ and $\boldsymbol{c_{ij}}$ quantify shared (family) and unique (individual) age effects, respectively. For this extended twin model, the methylation variance equals $\sigma^2_{tot} = \sigma^2_b + \sigma^2_e + age^2 * (\sigma^2_a + \sigma^2_c)$, and when the ages of the two twins in the twin-pair are approximately equal, then the correlation between the two methylation values within a twin-pair is given by $\rho = (\sigma^2_b + age^2 * \sigma^2_a) / \sigma^2_{tot}$. In that case, the within-pair difference $\boldsymbol{y_{i1} - y_{i2}}$ for this extended model has a normal distribution with mean zero and variance $2(\sigma^2_e + age^2 * \sigma^2_c)$. So for the extended model, both total variance, and the variance of the within-pair differences increase quadratically with age.

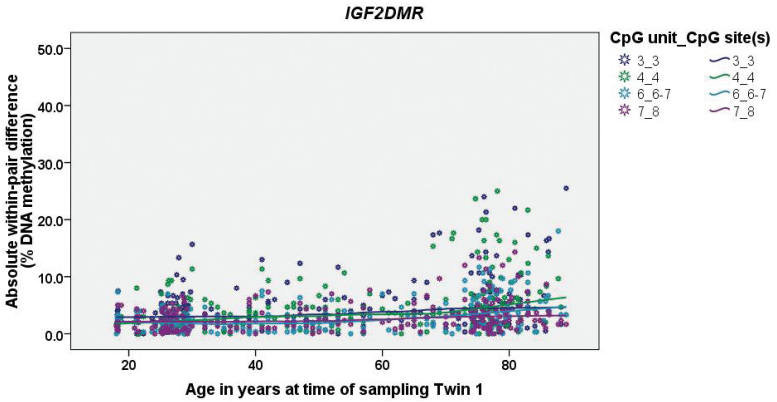
Supplementary Figures



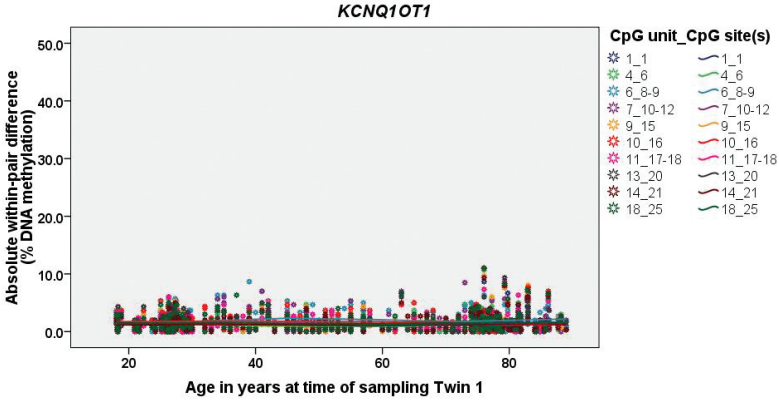
Supplementary Figure S4.1: The Absolute within MZ twin pair difference in % DNA methylation (y-axis) plotted against age (x-axis) per CpG unit of the *LEPTIN* locus. Each dot represents the absolute within pair difference of one MZ pair at one CpG unit. The different CpG units are represented by different colours. The lines are a loess lines that show the average trend for the change in methylation discordance at a CpG site with age.



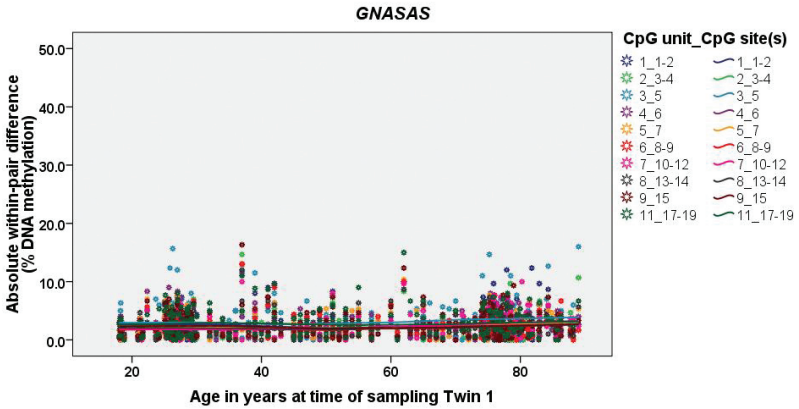
Supplementary Figure S4.2: The Absolute within MZ twin pair difference in % DNA methylation (y-axis) plotted against age (x-axis) per CpG unit of the *CRH* locus. Each dot represents the absolute within pair difference of one MZ pair at one CpG unit. The different CpG units are represented by different colours. The lines are a loess lines that show the average trend for the change in methylation discordance at a CpG site with age.



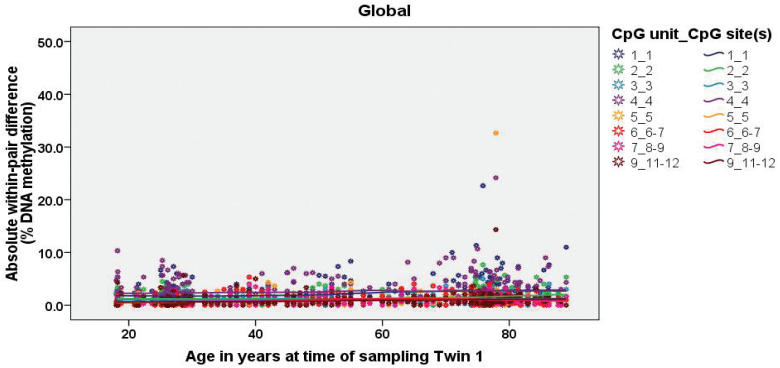
Supplementary Figure S4.3: The Absolute within MZ twin pair difference in % DNA methylation (y-axis) plotted against age (x-axis) per CpG unit of the *IGF2DMR* locus. Each dot represents the absolute within pair difference of one MZ pair at one CpG unit. The different CpG units are represented by different colours. The lines are a loess lines that show the average trend for the change in methylation discordance at a CpG site with age.



Supplementary Figure S4.4: The Absolute within MZ twin pair difference in % DNA methylation (y-axis) plotted against age (x-axis) per CpG unit of the *KCNQ1OT1* locus. Each dot represents the absolute within pair difference of one MZ pair at one CpG unit. The different CpG units are represented by different colours. The lines are a loess lines that show the average trend for the change in methylation discordance at a CpG site with age.



Supplementary Figure S4.5: The Absolute within MZ twin pair difference in % DNA methylation (y-axis) plotted against age (x-axis) per CpG unit of the *GNASAS* locus. Each dot represents the absolute within pair difference of one MZ pair at one CpG unit. The different CpG units are represented by different colours. The lines are a loess lines that show the average trend for the change in methylation discordance at a CpG site with age.



Supplementary Figure S4.6: The Absolute within MZ twin pair difference in % global DNA methylation (y-axis) plotted against age (x-axis) per CpG unit. Each dot represents the absolute within pair difference of one MZ pair at one CpG unit. The different CpG units are represented by different colours. The lines are a loess lines that show the average trend for the change in methylation discordance at a CpG site with age.

Supplementary tables

Table S4.1A: primers used for the BS PCR

Locus (assay)[a]	Forward primer	Reverse primer
ABCA1	ATTTTATTGGTGTTTTTGGTTGT	ATCAAAACCTATACTCTCCCTCCTC
CRH	TGGTTGTTGTTTTTTTGGTAGG	AATTTCTCCACTCCAAAACCTAAA
GNASAS	GTAATTTGTGGTATGAG-GAAGAGTGA	TAAATAACCCAACTAAATCCCAACA
IGF2_pter	GTTGTGTGTTTAGTGGTTTTT-GTTG	AAAAAATTTACCTA-AAAAAAACTTCCC
IGF2_DMR	TGGATAGGAGATTGAG-GAGAAA	AAACCCCAACAAAAACCACT
IGF2_qter	GATGAGGTTTTTTTATTTG-TAGGGG	AAAACCAAAATCCTAACAACTACCC
INS	GTTTTGAGGAAGAGGTGTTGA	ACCTAAAATCCAACCACCCTAA
KCNQ1OT1	TTTGGTAGGATTTTGTTGAG-GAGTTTT	CTCACACCCAACCAATACCTCATAC
LEP	GTTTTTGGAGGGATATTA-AGGATTT	CTACCAAAAAAAACCAACAAAAAAA
LINE1	GTGTGAGGTGTTAGTGTGTTTT-GTT	ATATCCCACACCTAACTCAAAAAAT
10-mer tag	AGGAAGAGAG + primer	
T7-tag		CAGTAATACGACTCACTATAGGGAGA-AGGCT + primer

Table S4.1B: CpG sites per CpG unit of each assay and assay call rates (CR) after quality control for the two phases of this study

		Young Dutch and old Danish MZ pairs	Middle-aged and old Dutch MZ pairs		Follow-up Danish MZ pairs	
Locus (assay)[a]	CpG sites analyzed[b]	CR (%)	CpG sites analyzed[b]	CR (%)	CpG sites analyzed[b]	CR (%)
ABCA1[c]	**1[c],** 2*, **3-4[c], 5[c], 6-9[c], 15-16, 17-18, 19-21, 24, 25**	99.1[c]	NA		NA	
CRH	**1, 2,** 3*, 4*, 8*, **9, 10**	93.4	Same CpG sites	89.7	Same CpG sites	88.5
GNASAS	**1-2, 3-4, 5, 6, 7, 8-9, 10-12, 13-14, 15, 17-19**	97.4	Same CpG sites	98.2	Same CpG sites	98.9
IGF2_pter	**1, 2, 3, 4, 5**	99.2	NA		NA	
IGF2_DMR	3, 4, 6-7, 8	99.7	Same CpG sites	99.1	Same CpG sites	99.3
IGF2_qter	1*, **2,** 7*, **8, 9,** 11*, **12-13**	76.5	NA		NA	
INS	**2, 3, 4, 5, 6**	98.9	NA		NA	
KCNQ1OT1	**1,** 2*, **6, 8-9, 10-12, 15, 16, 17-18, 20, 21, 25**	100	Same CpG sites	98.7	Same CpG sites	98.7
LEP	**1, 8, 16-17, 19-21, 22, 25**, 28*	98.2	Same CpG sites	95.7	Same CpG sites	94.1
LINE1	1, 2, 3, 4, 5, 6-7, 8-9, 11-12	99.5	Same CpG sites	96.9	Same CpG sites	94.6

a: Loci are ordered alphabetically in both tables

b: CpG site number is counted from the forward primer onward, CpGs that passed quality control are given in bold, CpGs that failed are marked by an asterix (*)

c: The average call rate of the 4 methylated CPG units at the 5' end of the amplicon (underlined) is given, the call rate for all CpG units was 99.5 %

Table S4.2: Significance of the test for interaction between country of origin and the observed age related epigenetic effects

	$P_{interaction}$[b]	
Locus (assay)[a]	**Inter- individual variation**	**Within-pair discordance**
Global	.739	.401
KCNQ1OT1	.444	.929
GNASAS	.182	.163
IGF2DMR	.465	.739
LEP	.098	.423
CRH	.285	.385

a: Loci are ordered from top to bottom as in Figure 4.1B from left to right

b: The test was done on a population of the old Dutch MZ pairs and old Danish MZ pairs

Table S4.3: Influence of percentage neutrophils on methylation variation and discordance in young MZ twins

	Inter-individual variation in DNA methylation		**Within-pair methylation discordance**	
Locus (assay)[a]	**% Variation explained[b]**	**p-value[c]**	**% Discordance explained[d]**	**p-value[c]**
Global	0.2	.464	1.0	.101
KC-NQ1OT1	0.1	.299	0.8	.519
GNASAS	0.2	.227	1.3	.164
ABCA1	0.1	.365	11.4	$3.7 *10^{-04}$
INS	0.6	.112	0.4	.157
IGF2DMR	0.3	.546	4.8	.008
IGF2_qter	2.8	.002	3.0	.045
IGF2_pter	1.9	.029	1.8	.062
LEP	13.7	$1.2 *10^{-10}$	19.4	$1.0 *10^{-06}$
CRH	0.3	.546	0.6	.762

a: Loci are ordered from top to bottom as in Figure 4.1A from left to right

b: Percentage of variation attributable to percentage neutrophil assessed from change in residual variance

c: Two sided p-value for the effect of neutrophil percentage in the appropriate nested linear mixed model

d: Percentage of twin discordance attributable to neutrophil percentage assessed from change in residual variance

Table S4.4: Influence of percentage neutrophils on age related changes in methylation variation and discordance

	Inter-individual variation in DNA methylation		Within-pair methylation discordance	
Locus (assay)[a]	**% Age related variation[b]**	**p-value[c]**	**% Age related discordance[d]**	p-value[c]
Global	10.0	.006	8.7	$8.2 * 10^{-04}$
KCNQ1OT1	0.0	.935	0.0	.794
GNASAS	1.2	.048	0.6	.578
IGF2DMR	3.8	.149	5.1	.132
LEP	9.7	$4.1 * 10^{-18}$	15.1	$4.2 * 10^{-12}$
CRH	0.5	.462	0.5	.631

a: Loci are ordered from top to bottom as in Figure 4.1B from left to right

b: Percentage of age related variation attributable to neutrophil percentage assessed from change in random effect estimate of age

c: Two sided p-value for the effect of neutrophil percentage in the appropriate nested linear mixed model

d: Percentage of age related twin discordance attributable to neutrophil discordance assessed from change in random effect estimate of age

Table S4.5: Mean (SD) DNA methylation and longitudinal change in inter-individual variation in old MZ pairs

	Mean (SD) in % DNA methylation[b]			Inter-individual variation	
Locus[a]	**Baseline**	**10 year follow-up**	**P_{Mean}[c]**	**Proportional increase[d]**	**$P_{variation}$[e]**
Global	59.3 (2.9)	62.0 (3.0)	.003	3.7 %	.400
KCNQ1OT1	30.1 (4.6)	31.7 (4.9)	.271	4.6 %	.999
GNASAS	40.7 (3.2)	42.6 (3.2)	.056	2.6 %	.442
IGF2DMR	54.2 (9.8)	46.0 (10.2)	.008	4.1 %	.301
LEP	31.5 (13.8)	34.0 (14.8)	.575	7.2 %	.046
CRH	67.3 (12.4)	68.2 (12.4)	.827	0.4 %	.344

a: Loci are ordered from top to bottom as in Figure 4.1C from left to right

b: DNA methylation and SD are adjusted for correlation between the CpG units, family relations, age, and sex

c: Two-sided p-value for testing of group means with a linear mixed model accounting for CpG units, sex, age, and family relations

d: The increase in SD during the follow-up period is given in proportion (percentage) to the SD at baseline

e: One-sided p-value for testing group variation with a Z-test on the estimates of variance from the same linear mixed model

Table S4.6: Significance test for the increase in total, familial and individual variation in DNA methylation

Locus[a]	Total variation[b]	Familial variation[b]	Individual variation[b]
Global	.004	.077	.003
KCNQ1OT1	$1.8 * 10^{-15}$	.999	$1.8 * 10^{-15}$
GNASAS	$1.2 * 10^{-05}$	.378	$1.5 * 10^{-11}$
IGF2DMR	.002	.077	$3.2 * 10^{-04}$
LEP	$3.5 * 10^{-05}$	.071	$4.2 * 10^{-10}$
CRH	$8.7 * 10^{-04}$	.007	.055

a: Loci are ordered from top to bottom as in Figure 4.2 from top left to bottom right

b: One-sided p-value for testing age related increase in variation with a Z-test on the corresponding estimates of variance from the same linear mixed model, effect sizes are shown in Figure 4.2

Table S4.7: Homogeneity of variance test for the within-pair methylation differences across the age groups per CpG unit

Locus_CpG unit	CpG site(s)	Up to 25 years old[a]			26 to 50 years old[a]			50 to 75 years old[a]			Over 75 years old[a]			P Levene's[c]
		SD[b]	Min[b]	Max[b]	SD[b]	Min[b]	Max[b]	SD[b]	Min[b]	Max[b]	SD[b]	Min[b]	Max[b]	
Global_1	1	2.8	-6.7	5.7	2.5	-6.2	7.3	5.0	-11.3	22.7	4.1	-9.0	11.0	.003
Global_2	2	1.8	-3.7	4.5	1.6	-4.0	4.0	2.2	-5.0	7.7	2.4	-5.7	5.3	.024
Global_3	3	1.2	-2.3	2.3	1.3	-3.3	2.7	1.6	-3.7	5.0	1.6	-4.0	3.7	.119
Global_4	4	3.9	-5.7	10.3	3.2	-6.0	7.0	4.2	-9.0	10.7	4.7	-7.0	24.2	.546
Global_5	5	1.1	-3.3	2.3	1.2	-2.7	4.3	1.6	-3.0	4.3	4.6	-3.2	32.7	.190
Global_6	6-7	1.0	-2.3	2.0	1.3	-5.3	3.0	1.8	-4.0	3.7	1.4	-3.7	4.0	.006
Global_7	8-9	0.8	-1.0	1.7	1.2	-4.0	2.3	1.5	-3.3	3.3	1.3	-3.3	3.0	.048
Global_9	11-12	1.3	-1.7	4.3	1.4	-5.7	2.7	1.4	-2.7	3.7	2.3	-14.3	3.0	.386
*KCNQ1OT1*_1	1	1.6	-3.3	3.3	2.1	-5.0	4.7	2.0	-6.7	4.7	2.8	-7.7	8.7	.057
*KCNQ1OT1*_4	6	1.9	-3.7	3.7	2.0	-5.7	5.7	2.0	-5.0	4.7	2.8	-8.3	10.7	.439
*KCNQ1OT1*_6	8-9	1.7	-3.0	3.7	3.0	-8.7	6.0	2.6	-6.3	5.0	2.8	-4.7	7.3	.072
*KCNQ1OT1*_7	10-12	1.5	-2.3	2.7	2.4	-7.0	5.3	2.4	-8.5	4.7	2.9	-5.7	11.0	.099
*KCNQ1OT1*_9	15	1.5	-2.3	3.7	1.9	-4.7	5.3	2.0	-5.0	4.3	3.0	-7.7	9.7	.048
*KCNQ1OT1*_10	16	2.2	-3.3	5.3	2.2	-4.7	5.0	2.2	-5.0	4.7	3.1	-8.0	11.0	.717
*KCNQ1OT1*_11	17-18	1.6	-3.3	3.3	2.2	-6.0	5.3	1.8	-5.0	3.7	2.6	-8.7	7.3	.240
*KCNQ1OT1*_13	20	1.8	-4.3	4.3	2.0	-5.0	4.7	1.8	-6.0	3.3	2.5	-9.3	6.0	.549
*KCNQ1OT1*_14	21	1.3	-2.3	2.7	1.7	-4.0	5.0	1.9	-5.0	4.3	2.8	-8.0	9.3	.023
*KCNQ1OT1*_18	25	1.8	-3.0	4.7	2.2	-6.3	4.7	2.1	-6.7	4.3	3.1	-8.3	11.0	.089
*GNASAS*_1	1-2	3.0	-6.0	4.7	4.0	-13.0	7.3	3.3	-10.0	9.7	4.1	-12.3	9.7	.399
*GNASAS*_2	3-4	2.6	-5.3	5.0	3.3	-14.7	6.0	3.1	-8.3	6.7	3.8	-9.7	10.7	.280
*GNASAS*_3	5	4.3	-7.0	12.3	4.8	-15.7	7.0	4.5	-14.7	7.7	5.7	-16.0	7.7	.357
*GNASAS*_4	6	3.8	-8.3	9.0	3.5	-11.3	7.0	3.1	-10.0	6.7	3.6	-6.3	7.3	.357
*GNASAS*_5	7	3.2	-7.0	6.0	3.5	-13.0	6.3	3.4	-10.3	7.7	3.4	-6.7	6.0	.927
*GNASAS*_6	8-9	2.6	-6.0	6.0	3.2	-13.0	5.3	3.0	-10.0	6.0	3.2	-6.3	7.7	.507
*GNASAS*_7	10-12	2.6	-6.0	6.0	3.3	-10.0	6.0	3.3	-9.7	7.0	3.8	-10.0	6.3	.146
*GNASAS*_8	13-14	2.5	-6.0	6.3	3.3	-12.0	6.7	3.1	-8.7	7.3	2.9	-6.0	5.7	.455
*GNASAS*_9	15	3.2	-6.7	7.0	3.7	-16.3	6.0	3.6	-12.3	7.3	3.4	-7.7	9.3	.910
*GNASAS*_11	17-19	3.0	-7.7	5.7	4.1	-11.0	8.0	4.1	-15.0	7.3	3.7	-7.0	8.0	.335
*IGF2DMR*_3	3	3.5	-7.5	5.0	5.0	-15.7	12.3	6.6	-18.2	10.7	9.6	-24.0	25.5	$6.0*10^{-05}$

Table S4.7: (continued) Homogeneity of variance test for the within-pair methylation differences across the age groups per CpG unit

Locus_CpG unit	CpG site(s)	Up to 25 years old[a]			26 to 50 years old[a]			50 to 75 years old[a]			Over 75 years old[a]			P Levene's[c]
		SD[b]	Min[b]	Max[b]	SD[b]	Min[b]	Max[b]	SD[b]	Min[b]	Max[b]	SD[b]	Min[b]	Max[b]	
IGF2DMR_4	4	2.9	-8.0	7.0	4.4	-11.3	9.7	7.1	-23.7	16.7	9.5	-25.0	21.7	$1.6*10^{-07}$
IGF2DMR_6	6-7	2.9	-7.3	5.0	2.6	-7.0	7.5	4.1	-10.7	11.7	5.9	-18.0	10.7	$3.3*10^{-07}$
IGF2DMR_7	8	2.6	-5.0	4.3	3.0	-6.7	7.3	4.3	-12.0	13.3	5.1	-14.3	12.0	$9.6*10^{-04}$
LEP_1	1	4.0	-8.0	8.7	4.9	-17.3	11.7	9.1	-22.3	30.0	10.6	-23.3	42.3	$6.8*10^{-05}$
LEP_3	8	2.8	-5.0	5.3	2.7	-5.0	10.7	7.0	-36.3	15.0	6.5	-17.3	16.3	$7.3*10^{-06}$
LEP_8	16-17	2.5	-5.0	5.0	2.7	-7.0	7.0	4.9	-11.3	14.0	5.3	-15.3	18.0	$4.9*10^{-04}$
LEP_10	19-21	3.7	-8.0	6.3	3.4	-8.0	7.7	4.3	-19.0	7.3	6.4	-26.0	11.2	.005
LEP_11	22	4.9	-8.0	10.0	5.3	-15.7	17.3	8.2	-24.3	21.3	10.1	-34.7	29.7	.002
LEP_13	25	6.8	-12.0	13.7	6.0	-12.3	14.7	10.2	-30.3	33.0	11.0	-37.7	16.3	.008
CRH_1	1	5.5	-7.7	17.7	6.8	-12.5	22.8	13.1	-46.3	32.5	10.1	-16.0	34.7	.003
CRH_2	2	3.5	-7.3	9.0	5.6	-19.7	21.3	4.5	-9.5	12.0	8.0	-29.7	14.0	.012
CRH_9	9	6.4	-12.0	15.0	8.6	-32.8	25.3	11.7	-27.0	28.3	12.2	-26.7	36.7	.015
CRH_10	10	5.7	-11.3	14.7	7.5	-22.0	20.3	15.3	-36.0	38.5	12.4	-29.7	26.0	$2.4*10^{-04}$

a: Population sizes: ≤ 25 yrs, n = 30 pairs; 26 – 50 yrs, n = 78 pairs; 51 – 75 yrs, n = 56 pairs; ≥ 76 yrs, n = 54 pairs

b: Standard deviation of within pair difference (SD) minimum (Min) and maximum (Max) differences are given in % DNA methylation

c: Two-sided p-value from Levene's test for homogeneity of variances across the age groups

Chapter 5

Hypermethylation at loci sensitive to the prenatal environment is associated with increased incidence of myocardial infarction

R.P. Talens[1], J.W. Jukema[2,3,4,5], S. Trompet[2], D. Kremer[1], R.G.J. Westendorp[3,6], L.H. Lumey[7], N. Sattar[8], H. Putter[9], P.E. Slagboom[1,3], B.T. Heijmans[1,3], on behalf of the PROSPER group

1. Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
2. Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands
3. Netherlands Consortium for Healthy Ageing, The Netherlands
4. Interuniversity Cardiology Institute of the Netherlands (ICIN), Utrecht, The Netherlands
5. Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands
6. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden. The Netherlands
7. Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA
8. Institute of Cardiovascular and Medical sciences, BHF Glasgow Cardiovascular Research Center, University of Glasgow, Glasgow, United Kingdom
9. Department of Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands.

Published in International Journal of Epidemiology, Volume 41, Issue 1, pages 106 – 115, 2012

Abstract

Background
Human epidemiological studies suggest that small size at birth and food deprivation during gestation confer an excess risk of coronary heart diseases (CHD) in adulthood, frequently in a sex-specific manner. Prior epigenetic studies indicate that such prenatal conditions are marked by persistent, sometimes sex-specific changes in DNA methylation. Here, we investigated the association between DNA methylation and myocardial infarction (MI) at six loci sensitive to prenatal nutrition, anticipating potential sex-specificity.

Method
Within the placebo group of the PROSPER trial on Pravastatin and the risk of CHD, we compared all individuals who were event free at baseline and developed MI during 3 years follow-up (n = 122) with a similar sized control group. Methylation at *IL10*, *LEP*, *ABCA1*, *IGF2*, *INS*, and *GNASAS* was measured in DNA extracted from leukocytes using mass spectrometry.

Results
DNA methylation at *GNASAS* was modestly higher in MI cases compared with controls (p=0.030). A significant sex interaction was observed for *INS* (p = 0.014) and *GNASAS* (p = 0.031). Higher DNA methylation at these loci was associated with MI among women (*INS*: + 2.5 %, p = 0.002; *GNASAS*: + 4.2 %, p = 0.001). Hypermethylation at one locus and at both loci was associated with odds ratios of 2.8 and 8.6, respectively ($P_{trend} = 3.0 * 10^{-04}$). No association was observed among men.

Conclusions
The risk of MI in women is associated with DNA methylation marks at specific loci previously shown to be sensitive to prenatal conditions. This observation may reflect a developmental component of MI.

Introduction

The developmental origins of health and disease hypothesis poses that conditions during early development influence the risk for developing coronary heart diseases (CHD) and other chronic diseases in later life [209,210]. In human studies, this hypothesis is being investigated using various proxies for developmental adversity. Small size at birth, for example, is linked to a consistent albeit modest increase in risk of type 2 diabetes mellitus [166] and CHD [9] in later life. Prenatal exposure to the Dutch Famine at the end of World War II is associated with an adverse CHD risk profile in adulthood in individuals who had a normal size at birth [167,211]. These proxies, however, generally lack specificity. Small size at birth can be caused by many factors and cannot distinguish between specific prenatal conditions and their timing. Studies of historical famines can be relatively precise for the type and the timing of prenatal conditions but provide limited information for exposures at the individual level. Moreover, data on these proxies is rarely available in epidemiological studies of CHD. The molecular modifications underlying epigenetic regulation are an alternative resource for markers of the prenatal environment that could meet these limitations.

Epigenetic mechanisms maintain heritable variation in gene expression potential that is not caused by variation in the DNA sequence [29]. The molecular basis includes the methylation of cytosines in CpG dinucleotides, which together with histone modifications, and non-coding RNAs, influences the accessibility of DNA to the transcriptional machinery [23]. DNA methylation is recognized as the most stable epigenetic mark [25]and can be measured on DNA samples that are commonly available in biobanks [195].

Although generally stable, the prenatal environment can induce persistent changes in DNA methylation. Crucially, these changes can be specific to the type of exposure, its timing during gestation and the sex of the exposed individual. In sheep, maternal methyl-donor deficiency from preconception to preimplantation resulted in sex-specific DNA methylation changes throughout the genome in the adult offspring [184]. In rats, prenatal dietary protein restriction [212] and prenatal caloric restriction [213] had opposite effects on DNA methylation at the same candidate

Table 5.1: Epigenetic properties of the six methylation assays and their associated candidate loci

	Prenatal famine (14, 15)						
Locus (alias)	**Peri[a]**	**Late[b]**	**Sex specific**	**mTFBS[c]**	**Epigenetic feature (10)**	**Control of gene**	**Gene function**
IL10	↑	=	n	Confirmed	Proximal promoter	*IL10* [136]	Anti-inflammation
LEP	↑[d]	↑[d]	y (♂)	Confirmed	Proximal promoter	*LEPTIN* [141]	Metabolism
ABCA1	↑	NT	n	Confirmed	Promoter CpG island shore	*ABCA1* [143]	Cholesterol transport
IGF2DMR	↓	=	n	Putative	Imprinted DMR	*IGF2* [133]	Growth and development
INS (*INSIGF*)	↓[d]	=	y (♂)	Putative	Proximal promoter	*INSULIN* [200] / *INS-IGF* [180]	Glucose metabolism / Fetal growth
GNASAS (*NESPAS*)	↑[d]	↓	y (♀)	Confirmed	Imprinting control DMR	*Gs, ALPHA SUBUNIT* [200]	cAMP-dependent pathway

a: Exposure to famine in the period around conception

b: Exposure to famine in late gestational period (last trimester)

c: Methylation-sensitive transcription factor-binding site

d: The association between DNA methylation and prenatal famine was sex specific

NT: The association between DNA methylation and famine was not tested

loci illustrating that methylation marks at specific loci can be meta-stable and that the direction of epigenetic changes may be exposure-specific [214]. Similar mechanisms apply to humans. We reported timing and sex-specific changes in DNA methylation at candidate loci sixty years after exposure to the Dutch Famine [14,135]. The relevance of sex-specific effects is highlighted by the observation that the association of prenatal famine exposure with an adverse cardiovascular risk profile is predominantly present among women [167,169]. In addition, periconceptional exposure to famine was associated with lower DNA methylation at the *IGF2* locus whereas periconceptional folic acid supplementation was associated with higher DNA methylation at this locus [120]. Other prenatal conditions that have been associated with

changes in DNA methylation in humans include intrauterine growth restriction [215] and maternal tobacco smoking during pregnancy [216]. In line with these data, it has been proposed that the developmental origins of CHD may eventually be investigated with epigenetic signatures that are specific for both timing and type of adverse prenatal conditions [69,106].

In this study, we investigated the association of DNA methylation with myocardial infarction (MI) during a 3.2 years follow-up period. To this end, DNA methylation was measured at six loci implicated in metabolic and cardiovascular diseases, namely *IL10*, *LEP*, *ABCA1*, *IGF2*, *INS* (alternate symbol *INSIGF*), and *GNASAS* (alternate symbol *NESPAS*). These methylation marks cover the genomic region that contains features with a potential for epigenetic regulation as observed in human, animal, or cell culture experiments [133,136,141,143,200,201] (Table 5.1). DNA methylation at these loci was previously found to mark prenatal environmental conditions [14,120,135]. Analyses took into account potential sex-specificity since prenatal adversity was reported to be sex-specifically associated with both changes in DNA methylation [135] and with prevalence of metabolic risk factors for CHD [167,169,217].

Methods and materials

Study population

This study was based in the ongoing PROspective Study of Pravastatin in the Elderly at Risk (PROSPER) trial. Details of the design of the PROSPER study were described previously [118,218]. Briefly: PROSPER is a prospective multicenter randomized placebo-controlled clinical trial to asses if treatment with pravastatin could reduce risk of major vascular events in the elderly. Between December 1997 and May 1999, subjects were screened and enrolled in Glasgow (Scotland), Cork (Ireland), and Leiden (the Netherlands). For the PROSPER study men and women aged 70 – 82 years were recruited if they had pre-existing vascular disease or were at a moderately increased risk due to smoking, hypertension, or diabetes. A total number of 5804 subjects were randomly assigned to either Pravastatin (n = 2891) or placebo (n = 2913) treatment. Mean follow-up period was 3.2 years after randomized assignment to Pravastatin or placebo treatment.

All endpoints of the study were adjudicated by a study endpoint committee. In the current study, MI includes the definitions of definite fatal and nonfatal myocardial infarction as devised for the PROSPER trial [118]. Further details on the measurement and definition of baseline characteristics are described elsewhere [118,218]. DNA was extracted from blood drawn before the placebo run-in medication was issued. Samples were centrifuged after lysis of erythrocytes, the supernatant was removed, and DNA was extracted from the pellet, consisting of the leukocytes, using the salting out method.

Eligible individuals for this study were allocated to placebo treatment to exclude an influence of Pravastatin treatment and were free of CHD events at baseline (n = 1654). The remaining risk factors for which the study population was selected (smoking, hypertension or diabetes) are common at older age and the individuals can be expected to be largely representative of their age group. A further inclusion criterion for this study was availability of sufficient DNA (n = 1375). Controls were selected from those alive at the end of the follow-up period who did not develop CHD, stroke or cancer (n = 1248), cases suffered an MI during the follow-up period (n = 122). The study focused on a single end-point to reduce heterogeneity. A case-control study was designed to compare all eligible cases to a representative control group of similar size. Assuming that the differences in DNA methylation between cases and controls are similar to those we previously observed between individuals prenatally exposed to famine [14,135] or periconceptional folic acid [120] (mean difference expressed in standard deviations = 0.35), the estimated power of our study is 0.78 (122 cases, 122 controls; $\alpha = 0.05$; two-sample t-test). Because this would leave space for 4 additional samples on the PCR-plates, 126 individuals were included in the control group. To select the controls, we used block-random selection in order to obtain a similar sex distribution in the controls as in the cases (females 48.4 %). From the group of eligible controls (females 60.9 %), 302 females were randomly removed. Next, from the remaining group (n = 946), 126 individuals (females 49.2 %) were randomly selected to be the control group, which was similar in baseline characteristics as the whole group of eligible controls.

The institutional ethics review boards of all study centers approved the protocol, and all participants gave written

informed consent. The protocol was consistent with the Declaration of Helsinki.

DNA methylation

DNA methylation was measured using a mass spectrometry-based method (Epityper version 1.05, Sequenom, San Diego, CA, USA). Previous studies showed the quantitative accuracy (R^2 duplicate measurements ≥ 0.98) and high concordance with clonal PCR bisulphite sequencing [14,151,152]. The design, features and measurement details of the assays to measure DNA methylation at *IL10*, *LEP*, *ABCA1*, *IGF2*, *INS* (*INSIGF*), and *GNASAS* (*NESPAS*) were described previously [195]. Briefly: bisulfite conversion of 0.5 µg of genomic DNA using the EZ 96-DNA methylation kit (Zymo Research, Orange, CA, USA) was followed by PCR amplification, fragmentation after reverse transcription and analysis on a mass spectrometer. Fragments that contain one or more CpG sites are called CpG units, methylation of which is assessed in a single measurement for all CpG sites on the fragment. Cases and controls were randomly divided over bisulfite conversion (3 x 96-well plates) and PCR amplification batches (2 x 384 well plates). For each individual, the six assays were amplified from the same bisulfite treated DNA. All methylation measurements were done in triplicate from the same bisulfite treated DNA.

The six assays contained a total of 59 CpG units, over which 96 CpG sites were distributed. After applying a stringent quality control procedure [195], 32 CpG units, containing 49 CpG sites remained (Table S5.1B). In total 23 CpG units (41 CpG sites) were removed since their fragment's mass overlapped or wass outside the detection range, potential confounding by single nucleotide polymorphisms both due to known (dbSNP build 28) or unknown genetic variation (R-module MassArray [153]), and 4 CpG units (6 CpG sites) were removed since they had success of less than two of the three replicate measurements, standard deviation (SD) greater than 10 % for the replicates, and success rate < 75 % after quality controls. The average call rate for the 32 CpG units was 94%.

Statistical analysis

The differences in baseline characteristics between the case and control group were assessed with a chi-square test for the qualitative variables, and an independent t-test for

the quantitative variables. For HDL-cholesterol, triglycerides, glucose and CRP, significance of the difference was assessed after natural logarithmic transformation.

We estimated the mean DNA methylation at the six loci and the standard error of the mean (SE) for cases and controls as previously described [135,195]. Linear mixed model were used to test for differences in methylation of the loci between cases and controls [135,195,219]. It uses all available methylation data per locus, i.e. methylation of multiple CpG units per locus, accounts for the correlation between methylation of CpG units within a locus, and using this correlation handles data missing at random. It further enables the inclusion of relevant adjustments on the raw data within the same model. The linear mixed model may be seen as an extension of a t-test, resulting in identical results when between-group methylation differences are assessed for a single CpG unit (in stead of all CpG units within a locus) and covariates are omitted. In the linear mixed model DNA methylation was entered as dependent variable. Individual was entered as a random effect. Incidence of MI, age, sex (when applicable), country of origin, PCR- and bisulphite batch, and CpG unit, were entered as fixed effects. The difference between case and control groups at a locus was expressed as percentage DNA methylation and as the proportion of the standard deviation (SD) of the adjusted mean methylation in the control group (estimated from the SE), i.e. the standardized difference (SD-unit).

To test for sex specificity of the association between DNA methylation and MI the interaction term MI*sex was entered as a fixed effect to the model. This was done in view of the previously observed sex-specificity of the association between prenatal exposure to famine and both DNA methylation at *LEP*, *INS*, and *GNASAS* [135] (Table 5.1) and cardiovascular risk factors [167,169]. For loci showing a significant interaction between sex and MI, the study population was reanalyzed for men and women separately. Traditional CHD risk factors, that showed a difference ($p < 0.05$) between cases and controls, were added to the model as fixed effects, separately and simultaneously to obtain adjusted DNA methylation differences between cases and controls.

Logistic regression was used to calculate odds ratios (OR) for hypo– (i.e. lower than the median) or hypermethylation (i.e. higher than the median) at a locus and incidence of

MI. The median, being independent of the distribution of variation in DNA methylation, was chosen as cut-off to compare the analyses across the loci. To categorize DNA methylation at a locus as hypo- or hypermethylated, the mean methylation at a locus was calculated by averaging the methylation of the CpG units measured across the locus after imputing missing data. Imputation was performed using the linear mixed model, which exploits the correlation between methylation of neighboring CpG sites to achieve accurate imputation [14]. In the regression model, incidence of MI status was the dependent variable, and methylation status, age, sex (when applicable), country of origin, PCR- and bisulphite batch, and traditional risk factor (when applicable) were entered as independent variables. Odds ratios for hypo - and hypermethylation were calculated with methylation of a locus entered as categorical covariate (hypomethylation was the reference category). Loci were tested individually and combined; interaction between loci was tested by entering an interaction term $locus_1$*$locus_2$ as covariate. Statistical significance for combined effects was assessed using a test for trend. All P-values reported are two-sided. All analyses were performed using SPSS 16.0 (SPSS inc., Chicago, IL, USA).

Results

Our study was based in the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER) trial [118]. Eligible individuals for the current epigenetic study were selected from the placebo group and were free of CHD events at baseline (n = 1654). Individuals who developed MI during follow-up (n = 122) had a higher BMI, lower HDL-cholesterol, higher triglycerides, and higher CRP than a randomly selected control group (n=126; Table 5.2). Of note, the association of many CHD risk factors had a different distribution between the two sexes in this study population. Metabolic risk factors such as BMI, triglyceride levels, and fasting blood glucose were only associated with MI in women, whereas CRP levels and hypertension were only associated with MI in men (Table 5.2).

DNA methylation at *IL10*, *LEP*, *ABCA1*, *IGF2*, *INS*, and *GNASAS* was measured using a mass spectrometry-based method [152]. Linear mixed model were used to test for

Table 5.2: Baseline characteristics of the case and control group

	Selected individuals			Females			Males		
Characteristic[a]	**Controls (n = 126)**	**Cases (n = 122)**	**p-value**	**Controls (n = 62)**	**Cases (n = 57)**	**p-value**	**Controls (n = 64)**	**Cases (n = 65)**	**p-value**
Current smoker	48 (38 %)	43 (35 %)	.642	18 (29 %)	15 (26 %)	.741	30 (47 %)	28 (43 %)	.665
Diabetes	12 (10 %)	17 (14 %)	.280	4 (7 %)	9 (16 %)	.103	8 (12 %)	8 (12 %)	.974
Hypertension	86 (68 %)	96 (79 %)	.063	51 (82 %)	46 (81 %)	.827	35 (55 %)	50 (77 %)	.008
Age years	74.8 ± 3.1	75.4 ± 3.6	.209	75.1 ± 3.2	75.4 ± 3.5	.654	74.6 ± 2.9	75.3 ± 3.7	.184
BMI kg/m^2	26.3 ± 4.4	28.9 ± 4.2	.004	26.4 ± 4.7	28.5 ± 4.6	.016	26.3 ± 4.1	27.4 ± 3.8	.102
Total Cholesterol mmol/l	5.7 ± 0.97	5.5 ± 0.89	.116	6.2 ± 0.94	5.8 ± 0.96	.034	5.3 ± 0.78	5.3 ± 0.74	.887
LDL-chol mmol/l	3.8 ± 0.89	3.7 ± 0.78	.173	4.1 ± 0.92	3.8 ± 0.85	.052	3.5 ± 0.74	3.5 ± 0.69	.764
HDL-chol[b] mmol/l	1.3 (1.10 - 1.55)	1.2 (0.97 - 1.41)	.003	1.4 (1.17 - 1.68)	1.3 (1.06 - 1.50)	.016	1.2 (1.01 - 1.41)	1.1 (0.92 - 1.25)	.066
Triglyceride[b] mmol/l	1.3 (1.05 - 1.74)	1.5 (1.10 - 2.06)	.038	1.4 (1.13 - 1.86)	1.6 (1.34 - 2.19)	.048	1.3 (0.98 - 1.69)	1.4 (1.00 – 2.00)	.237
Glucose[b] mmol/l	4.9 (4.60 - 5.33)	5.1 (4.70 - 5.90)	.039	4.7 (4.50 - 5.30)	5.2 (4.60 - 5.95)	.031	5.0 (4.60 - 5.48)	5.0 (4.70 - 5.75)	.511
CRP[b] mg/l	2.9 (1.48 - 6.01)	4.5 (1.82 - 7.54)	.024	3.1 (1.49 - 6.10)	4.7 (1.50 - 8.16)	.519	2.7 (1.33 - 5.77)	4.4 (2.00 - 7.27)	.013

a: Qualitative characteristics are given as their count (%), quantitative characteristics as mean ± SD

b: These quantitative variables are given as median (inter quartile range), due to there skewed distribution

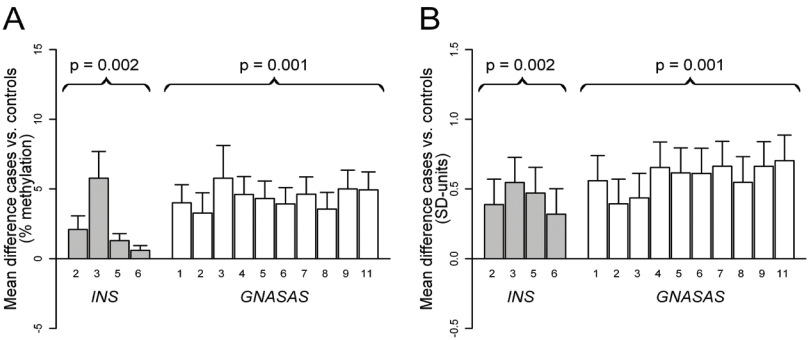


Figure 5.1: DNA hypermethylation is associated with risk for MI
DNA methylation difference between the female case - and control groups (y-axis) at each CpG unit (x-axis) of the loci *INS* (grey bars) and *GNASAS* (open bars). Differences were nominally significant for all CpG units when tested individually (P<0.05), except for CpG 6 of *INS* (**A**) Difference in percentage DNA methylation. (**B**) Difference in SD-units, a proportion of the standard deviation from the adjusted mean methylation in the control group. Bars represent the average difference, the whiskers represent the SE of the difference. Numbers under each bar are the CpG units, numbered from the forward primer onward. A positive difference indicates that the case group had a higher average DNA methylation, values are adjusted for bisulphite - and PCR batches, country and age at baseline.

differences in methylation of the loci between cases and controls [135,195,219]. DNA methylation was modestly higher in MI-cases at *GNASAS* (+ 1.8 % (= 0.26 SD-units), p = 0.030; Table 5.3) compared with the control group. No differences in DNA methylation were observed at the other loci. Sex-specificity of the associations was tested. An interaction between sex and MI was observed for *INS* ($P_{interaction}$ = 0.014) and *GNASAS* ($P_{interaction}$ = 0.031) (Table 5.3). Women and men were subsequently analyzed separately for these loci.

In women, DNA methylation at *INS* was higher in MI cases compared with controls (+ 2.5 % (= 0.59 SD-units), p = 0.002) (Table 5.4), which was independent of traditional risk factors (+ 2.6 % (=0.62 SD-units); p = 0.002). This difference was similar for all four CpG units at the locus, when expressed in SD-units (Figure 5.1B). DNA methylation at *GNASAS* was also higher in female cases compared with controls (+ 4.2 % (= 0.61 SD-units), p = 0.001), and was also independent of traditional risk factors (+3.5 % (=0.51 SD-units); p = 0.009). This difference was similar for all ten CpG units at this locus (Figure 5.1). In men, no differences in

Table 5.3: DNA methylation differences and incidence of MI

	Mean methylation[a] (SE)		Cases – Controls (SE)			
Locus	Controls	Cases	% Methyl-ation[b]	Standard-ized[c]	p-value[d]	$p_{sex\ interaction}$
IL10	27.6 (0.79)	29.2 (0.88)	+ 1.6 (1.11)	+ 0.18 (0.12)	0.148	0.232
LEP	29.2 (0.54)	29.7 (0.60)	+ 0.4 (0.75)	+ 0.07 (0.12)	0.558	0.435
ABCA1	14.8 (0.35)	15.0 (0.38)	+ 0.2 (0.47)	+ 0.04 (0.12)	0.752	0.093
IGF2	52.0 (0.52)	53.2 (0.58)	+ 1.2 (0.73)	+ 0.21 (0.13)	0.089	0.087
INS	75.7 (0.42)	76.8 (0.46)	+ 1.1 (0.58)	+ 0.23 (0.12)	0.063	0.014
GNASAS	49.0 (0.60)	50.9 (0.67)	+ 1.8 (0.84)	+ 0.27 (0.12)	0.030	0.031

a: Average DNA methylation at the locus, adjusted for batch effects, country, age at baseline, and sex

b: The difference is expressed in percentage DNA methylation

c: The standardized difference is expressed as a proportion of the standard deviation in the control group

d: Two sided p-value for the difference, adjusted for batch effects, country, age at baseline, and sex

DNA methylation were observed between cases and controls.

We assessed the association between hypermethylation at *INS* and at *GNASAS* with the risk of MI (Table 5.5). No interaction between the two loci was observed. Therefore, three categories were analyzed: hypomethylation at both loci (reference), hypermethylation at either *INS* or *GNASAS* and hypermethylation at both loci. In the whole study population, hypermethylation at one locus and at both loci was associated with odds ratios of 1.7 (95 % Confidence interval (95 % CI): 0.9 – 3.3) and 2.8 (95 % CI: 1.4 – 5.9), respectively ($p_{trend} = 0.006$). This association could be solely attributed to women among whom single and double hypermethylation were associated with odds ratios of 2.8 (95 % CI: 1.0 – 7.8) and 8.6 (95 % CI: 2.7 – 27.9), respectively ($p_{trend} = 3.1 * 10^{-4}$). Adjustment for traditional risk factors did not appreciably alter these associations ($OR_{single} = 3.1$ (95 % CI: 1.0 - 9.7); $OR_{double} = 11.1$ (95 % CI: 2,9 - 43.4), $P_{trend} = 5.2 * 10^{-4}$). In men, no association was observed.

Discussion

We report the association of a higher DNA methylation at the imprinted loci *INS* and *GNASAS* with the incidence

Table 5.4: DNA methylation differences and incidence of MI, per sex

Locus	Sex	Mean methylation[a] (SE)		Cases – Controls (SE)		
		Controls	Cases	% Methylation[b]	Standardized[c]	p-value[d]
INS	♂	76.0 (0.64)	75.7 (0.68)	− 0.4 (0.87)	− 0.07 (0.17)	0.659
	♀	75.5 (0.53)	77.9 (0.60)	+ 2.5 (0.76)	+ 0.59 (0.18)	0.002
GNASAS	♂	49.3 (0.83)	49.6 (0.89)	+ 0.2 (1.12)	+ 0.03 (0.17)	0.850
	♀	48.4 (0.89)	52.6 (1.01)	+ 4.2 (1.28)	+ 0.60 (0.18)	0.001

a: Average DNA methylation at the locus, adjusted for batch effects, country, and age at baseline

b: The difference is expressed in percentage DNA methylation

c: The standardized difference is expressed as a proportion of the standard deviation in the control group

d: Two sided p-value for the difference, adjusted for batch effects, country, and age at baseline

Table 5.5: Hyper-methylation at *INSIGF* and *GNASAS* and risk of MI

Methylation status *INSIGF* / *GNASAS*[a]	Whole group		♂		♀	
	n	OR (95% CI)	n	OR (95% CI)	n	OR (95% CI)
Hypo / Hypo	69	1	34	1	35	1
Hypo / Hyper or Hyper / Hypo	103	1.7 (0.9 – 3.3)	59	1.1 (0.4 – 2.7)	44	2.8 (1.0 – 7.8)
Hyper / Hyper	70	2.8 (1.4 – 5.9)	33	1.1 (0.4 – 3.1)	37	8.6 (2.7 – 27.9)
Test for trend (p_{trend})[b]		0.006		0.868		$3.1 * 10^{-4}$

a: Hypo-methylation is defined as below median, hyper-methylation as above median.

b: Two sided p-value, adjusted for batch effects, country, age at baseline, and sex where appropriate

of MI in women. Until now, only the association between various measures of genomic methylation and CHD have been investigated with varied outcomes [220,221]. Our results suggest that methylation at specific loci may be an epigenetic marker for the risk of CHD.

It remains to be determined which factors or mechanisms underlie the observed association. Our results may be linked

to the developmental origins hypothesis which predicts that individuals who go on to develop CHD were subjected to less favorable conditions in early life than those who do not [209,210,222]. *INS* and *GNASAS* are regulators of fetal growth. Methylation at *GNASAS* is crucial for normal imprinting throughout the *GNAS* locus, in which expression of the antagonists G protein α-subunit (Gαs) and its extra large isoform (XLαs) regulate fetal nutrient demand [201,223]. Methylation at the *INS* promoter establishes imprinting in extra-embryonic tissues, in which expression of the *INSIGF* transcript regulates placental development [180,224]. Further highlighting the putative link to the developmental origins hypothesis are the differences in DNA methylation at *INS* and *GNASAS* observed after prenatal exposure to famine that persist throughout life, are of the same magnitude and are sex-specific [135].

Our findings are compatible with a developmental component of MI, however the current study population lacks data on the potential prenatal factors involved, although it is certain that participants of the study population did not experience famine during gestation. Previous work indicated that epigenetically meta-stable loci like *INS* and *GNASAS* may be sensitive to various perturbations during early development [14,120,212,213]. Prenatal conditions that are associated with an adverse CHD risk profile in adulthood and that have been suggested to produce persistent epigenetic changes include maternal micronutrient deficiency [225], maternal smoking [226], and placental morphology [217]. Social-economic status may be another relevant but as yet less investigated factor related to prenatal conditions [8]. If persistent epigenetic signatures of specific prenatal conditions can be established in studies in which these prenatal conditions have been recorded [195,197], they can subsequently be used as surrogate markers to indirectly assess the association of such prenatal conditions to CHD risk in studies that, like our study, lack information on the prenatal environment [69,106].

The loci studied here were selected for their sensitivity to prenatal conditions and the influence of prenatal conditions on DNA methylation is well-established. However, the difference between female MI cases and controls may also have accumulated during aging as a result of a differential exposure to environmental factors over the life course [227].

If life course exposures indeed contributed to our findings, they appear to be unrelated to traditional CHD risk factors, since the associations between DNA methylation and MI remained after adjustment for these CHD risk factors. Similar to the studies on prenatal exposures, we measured DNA methylation on DNA extracted from leukocytes. We previously demonstrated that heterogeneity of the leukocyte population does not influence *INS* and *GNASAS* methylation in whole blood [195] and, furthermore, individuals with signs of inflammation were excluded from the current study [118]. Thus, it is unlikely that cellular heterogeneity contributed to our results.

Epigenetic changes were proposed to be mechanistically involved in CHD development [65], and it may be explored whether the current data can be viewed as supportive of this hypothesis. Higher DNA methylation at *INS* is associated with lower expression of the insulin gene in pancreatic β-cells [200]. Higher DNA methylation at *GNASAS* is associated with higher expression of Gαs in adipose tissue, the pituitary and the thyroid gland [201,223]. Gαs is a key component of the cAMP mediated intracellular signal transduction cascade [223]. Both cAMP and insulin signaling are implicated in obesity, insulin resistance and other signs of metabolic dysregulation [228], all of which are risk factors for developing CHD. The differences in DNA methylation that we report here, however, were observed in leukocytes, in which insulin is not expressed [200] and expression of Gαs is not controlled by DNA methylation at *GNASAS* [223]. Still, the methylation status of imprinted loci in leukocytes is generally thought to mark that of (internal) tissues that are directly involved in disease [30,195]. Also, it should be noted that the observed differences in percentage DNA methylation are small, although their magnitude is in line with other human studies [14,73,120,135,215,216]. In a rat study small changes in percentage DNA methylation were demonstrated to explain almost half the variance in gene expression [188]. Studies in non-peripheral tissues, incorporating a study design that addresses confounding and reverse causation [99], should be used to establish whether or not small but long-term changes in expression of these loci can influence pathways that are involved in CHD development.

In the current study we only found an association between DNA methylation and risk of MI in women, but not in men.

This may indicate sex-specificity of the association, similar to sex-specific epigenetic changes associated with prenatal exposure to the Dutch Famine [135]. In view of the study size and the characteristics of the PROSPER study, the absence of an association in men should be interpreted with caution. We observed sex-specific associations with MI for traditional risk factors that are known to be sex-independent in middle age. This may be related to the age range of the study subjects (70-82 years) and the exclusion of individuals with a history of CHD at baseline. The age is well above the average age of a first MI for men (63 years), but below the average age of a first MI for women (74 years) [229,230]. To definitely establish whether the associations of *INS* and *GNASAS* methylation are sex-specific, larger series of patients and controls that include younger age-ranges should be investigated.

It has been proposed that epigenetic marks that are sensitive to the prenatal environment can be used as biomarkers to investigate the developmental origins of health and disease hypothesis [69,106]. Our data, in combination with previous results on prenatal exposures [14,120,135] indicate the feasibility of this approach. The success of the approach will rely on the identification of epigenetic signatures that mark a specific prenatal condition, e.g. maternal micronutrient deficiency [225], maternal smoking [226], and placental morphology [217]. This will entail epigenome-wide screening in studies designed to investigate exposures during early development. These epigenetic signatures can subsequently be utilized as biomarkers for disease risk in epidemiological studies of CHD or other chronic diseases.

Acknowledgements

This work was performed as part of an ongoing collaboration of the PROSPER study group in the universities of Leiden, Glasgow and Cork. This work was supported by grants from the Netherlands Heart Foundation (#2006B083), the Interuniversity Cardiology Institute of the Netherlands (ICIN) and the Durrer Center for Cardiogenetic Research both Institutes of the Netherlands Royal Academy of Arts and Sciences (KNAW), an investigator initiated grant from Bristol-Myers Squibb (USA), the European Union–funded Network of Excellence LifeSpan (FP6/036894), the Center for Medical Systems Biology (CMSB), a center of excellence approved by the Netherlands Genomics Initiative (NGI) / Netherlands Organization for Scientific Research (NWO), the NGI/NWO-funded Netherlands Consortium for Healthy Ageing (NCHA) (#05060810) and the EU project HEALTH-F2-2007 223004 PHASE. J.W. Jukema is an established clinical investigator of the Netherlands Heart Foundation (#2001D032). The funders had no role in study design, data collection and analysis, decision to publish or the preparation of the manuscript.

Table S5.1A: Primers used in bisulfite PCR

Locus	Forward primer	Reverse primer
IL10	TGATTGGTTGAATATGAATTTTTGTAT	CACCCCCTCATTTTTACTTAAAAA
LEP	GTTTTTGGAGGGATATTAAGGATTT	CTACCAAAAAAAACCAA-CAAAAAAA
ABCA1	ATTTTATTGGTGTTTTTGGTTGT	ATCAAAACCTATACTCTCCCTCCTC
IGF2	TGGATAGGAGATTGAGGAGAAA	AAACCCCAACAAAAACCACT
INS	GTTTTGAGGAAGAGGTGTTGA	ACCTAAAATCCAACCACCCTAA
GNASAS	GTAATTTGTGGTATGAGGAAGAGTGA	TAAATAACCCAACTAAATCCCAACA
10-mer tag[a]	AGGAAGAGAG +	
T7 tag[a]		CAGTAATACGACTCAC-TATAGGGAGAAGGCT +

a: The primer is attached to the tag at the + site

Table S5.1B: CpG-sites per fragment of the loci that were analyzed for variation

Locus	UCSC coordinates[b]	CpG-sites analyzed[c]	Call rate[d]
IL10	chr1:206,946,011-206,946,339	1, 2-3	91%
LEP	chr7:127,881,054-127,881,410	1, 8, 16-17, 19-21, 22, 25, 28	95%
ABCA1	chr9:107,690,502-107,690,821	3-4, 5, 15-16, 17-18, 19-21, 25	92%
IGF2	chr11:2,169,459-2,169,796	3, 4, 8	94%
INS	chr11:2,182,336-2,182,640	2, 3, 5, 6	99%
GNASAS	chr20:57,425,815-57,426,108	1-2, 3-4, 5, 6, 7, 8-9, 10-12, 13-14, 15, 17-19	92%

b:UCSC genome browser coordinates from Feb. 2009 (GRCh37/hg19) Assembly

c: CpG-site number is counted from the forward primer onward, and given per CpG unit

d:Percentage of CpG unit methylation measurements that remained after quality control

Chapter 6

Genome Wide Differences and Similarities in DNA Methylation Between Internal Tissues and Blood

Rudolf P Talens,[1] Steffan D Bos,[1,2] Maurits RA Drijfhout van Hooff,[1] Judith VMG Bovée,[3] Ruud van der Breggen,[1] Nico Lakenberg,[1] Erik B van den Akker,[1,4] J Wouter Jukema,[2,5] P Eline Slagboom,[1,2] Ingrid Meulenbelt,[1,2] Jelle J. Goeman,[6] Bastiaan T. Heijmans[1,2]

1. Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
2. Netherlands Consortium for Healthy Ageing
3. Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands
4. Department of Mediamatics, Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands
5. Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands
6. Department of Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands

Data presented in this manuscript was augmented with additional data, reanalysed and published in Slieker et al., Epigenetics & Chromatin, Volume 6, article 26, pages 1 – 12, 2013.

Abstract

Background
Human studies on epigenetic variation and disease are usually restricted to measuring peripheral tissues, while many diseases manifest in internal tissues. Due to the extensive epigenetic remodeling involved in tissue development, it is unclear to what extent inter-individual epigenetic variation measured in peripheral tissues marks that of the internal tissues related to the disease.

Methodology & Principal Findings
To investigate epigenetic differences and similarities between blood and internal tissues we obtained samples from 6 individuals during post-mortem examination within 24 hours after death. Using the Illumina Infinium HumanMethylation450 BeadChip microarray We compared DNA methylation of blood with subcutaneous fat, skeletal muscle, visceral fat, liver, pancreas and spleen at 378,239 CpG sites distributed over the autosomal chromosomes. Principal component analysis on this data revealed that most differences in genome wide DNA methylation patterns were between tissues rather than individuals. This analysis also revealed that the methylation patterns in post-mortem blood samples were similar to the patterns in blood samples form 40 middle aged regular blood donors. The tissues showed a similar overall distribution of mean methylation levels and methylation variation at individual CpG sites. A quarter of the CpG sites measured showed no detectable methylation difference between blood and any of the internal tissues. Further, the methylation variation at more than 30% of CpG sites showed a strong correlation ($r \geq 0.75$) between blood and an internal tissue.
Most strong correlations were between blood and one internal tissue exclusively, although strong correlations between blood and several tissues were observed at 6092 CpG sites. searching for characteristics marking CpG sites with a strong tissue correlation, we found that these were unrelated to DNA methylation levels, variation, or differences between the tissues, nor did we observe an association with genetic sequence motifs, genome structure, or localization relative to genomic, epigenomic, or repetitive elements.

Conclusions
Variation in DNA methylation is correlated between blood and internal tissues at a subset of CpG sites, depending on both the genomic locus and the tissue. These results indicate the existence of loci at which epigenetic information in peripheral tissues marks that of internal tissues.

Introduction

A popular theme in research on the development of common age related diseases is its epigenetic component [17,65,68,231]. Epigenetic mechanisms regulate the local condensation of DNA thereby influencing the capacity of the transcription machinery to access a locus upon reception of the correct cues [21,23,29]. There are several correlated layers of epigenetic information, including post-transcriptional control by microRNAs, nucleosomal packaging, histone modifications and DNA methylation [22–25]. Epigenetic remodeling of cellular expression potential during embryogenesis is commonly thought to create the different cellular phenotypes from the single genotype of a zygote [182]. Epidemiological studies investigating the epigenetic component of common diseases mainly focus on DNA methylation since it can be reliably measured on DNA samples that are stored in a biobank, even after several decades [195].

Associations between locus specific DNA methylation and risk for rheumatoid arthritis [232], obesity [48] and myocardial infarction [233] have recently been reported. However, as bone [232], visceral fat [48] and myocardium [233] are unavailable for such studies, these associations were found in blood, which is readily available but not directly involved in the disease. Thus, the epigenetic contribution to these diseases remains unclear, since the relation between DNA methylation in blood and that of other tissues is not yet established. This is a pressing issue in epigenetic epidemiological research for most diseases. Clinical biobanks store DNA that was extracted from peripheral tissues, usually blood, occasionally buccal cells, and sporadically skin, subcutaneous fat, and skeletal muscle. This requires epigenetic studies to carefully consider the availability of tissues and the relevance of the information that can be obtained [98,99,202].

Documenting the loci at which DNA methylation measured in blood is a useful marker for inaccessible disease related tissues, will benefit interpreting the results of epigenetic studies, and may help even designing these studies if many such loci exist. DNA methylation differences between tissues have been reported at many loci [100–102,234]. Importantly though, despite the focus of these studies on tissue

differences, they also observed genomic areas with similar DNA methylation between the tissues within an individual [100,234]. Moreover, a study on candidate loci found a strong correlation between DNA methylation in blood and buccal swabs at 4 of the 8 loci investigated [195] and a genome wide study found loci with correlations between blood and brain cortex or cerebellum [234]. Together these results suggest that for a potentially subset of loci, DNA methylation in blood may mark the methylation status of another (disease related) tissue.

In this study we investigate epigenetic similarities between blood and several internal tissues. We obtained post mortem samples of blood, subcutaneous fat (SC fat), muscle, visceral fat (VS fat), liver, spleen, and pancreas from six individuals. We use the Illumina Infinium HumanMethylation450 BeadChip microarray (Illumina, San Diego, USA) to measure DNA methylation at 485,462 CpG sites distributed throughout the genome [85]. We focused on the 378,239 CpG sites for which the measurement was not influenced by known genetic variation. Principal component analysis (PCA) was used to investigate genome wide patterns of DNA methylation. At each of these 378,239 CpG sites we compared average DNA methylation and computed correlation coefficients between tissues. Our exploration of tissue correlations focused on 219,558 CpG sites which had multiple nearby CpG sites measured, using smoothed correlations over these CpG sites to indicate a more robust local effect [195].

Methods and materials

Subjects

The samples in this study were collected during post-mortem examination (within 24 hours after death) of 6 individuals (3 men) between 58 and 79 years old. From each individual we collected samples of 6 different tissues. Samples of blood, skeletal muscle, subcutaneous fat (SC fat) and visceral fat (VS fat) were collected from each individual, liver samples were collected from 5 individuals, pancreas from 4 individuals and spleen from 3 individuals. Autopsy case with tissue sample information is listed in Table 6.1. Blood was drawn from the carotid arteries, collected in BD

Table 6.1: Characteristics of selected individuals

Subject#	Gender	Age	PMI (h)	Cause of death	Tissues						
1	Female	61	7	Sepsis	B	F	V	M	L	S	
2	Male	58	12.5	Heart failure	B	F	V	M	L	S	
3	Male	64	9	Myocardial infarction	B	F	V	M	L		P
4	Female	65	15	Aortic aneurysm	B	F	V	M	L		P
5	Female	66	10	Liver cirrhosis	B	F	V	M		S	P
6	Male	79	10	Sepsis	B	F	V	M	L		P

PMI: Post Mortem Interval in hours

Tissues: B = Blood, F = Subcutaneous fat, V = Visceral fat, M = Muscle, L = Liver, S = Spleen, P = Pancreas

vacutainer EDTA blood collection tubes (Becton, Dickinson and Company, Franklin Lakes, NJ, USA), and treated according to standard protocol. The other tissues collected were dissected at the time of autopsy, glued to a piece of cork using Tissue-Tek OCT Compound (Sakura Finetek, Alphen aan de Rijn, Netherlands), snap-frozen in liquid nitrogen and stored at -80° C until further use. With a CM3000 Cryostat (Leica, Wetzlar, Germany) cryosections were made of the tissues. Two slices of 5 µm thick at either end of a tissue sample were stained with hematoxylin and eosin (H&E) to check tissue integrity and cell type heterogeneity, since each sample represent a mixture of the cell types normally found in that organ. All tissue from which DNA was isolated were histologically normal (Supplementary figure S6.1) and of similar consistency between the two HE-stained section of an individual, indicating an even heterogeneity throughout the tissue sample. Tissue heterogeneity was also similar between individuals, except for visceral fat where substantial differences in heterogeneity were observed between individuals (Supplementary figure S6.1).

DNA from blood was extracted from the buffy coat layer after centrifugation. Tissue in between the histological slides was cut into 20 µm thick slices, 30 to 40 of which were lysed overnight at 56° C in QIAGEN ATL buffer + proteinase K (Qiagen, Düsseldorf, Germany). DNA was extracted using a double Phenol : Chloroform : Isoamyl alcohol (25:24:1) (PCI) separation. The supernatant of the first separation, was treated with PCI once again to remove all residual traces of fatty acids, which was especially necessary for tissues with a

high lipid content. The remainder of the procedure followed the standard protocol.

All samples were obtained with approval of the pathology department of the LUMC and according to the 'Proper Secondary Use of Human Tissue' code of conduct by the 'Federatie van Medisch Wetenschappelijke Verenigingen'.

Genome wide DNA methylation

DNA methylation was measured at 485,462 CpG sites using the Illumina Infinium HumanMethylation450 BeadChip array (Illumina, San Diego, USA) according to manufacturer's protocol [85]. In short: about 700 ng of genomic DNA was bisulphite converted using the EZ-96 DNA Methylation Kit (Zymo Research, Orange County, USA). Each DNA sample was whole-genome amplified, and enzymatically fragmented. The DNA fragments were hybridized to locus-specific DNA oligomers linked to individual bead types. After single base pair extension, the array was fluorescently stained, scanned and the intensities of the non-methylated and methylated bead types were measured.

All samples of an individual were measured on the same chip. To check whether DNA methylation is affected by death, we compared DNA methylation of the post mortem blood samples with DNA from blood samples of living blood donors (35 – 40 years old), who represent DNA methylation in the healthy adult population [235]. This DNA was measured in two mixtures, one containing DNA of twenty men, the other of twenty women. Methylation of pooled DNA samples was recently demonstrated to accurately reflect the group average [236].

Quality control

Technical variability between chips was tested by measuring a duplo series of methylation titers from standardized non-methylated and methylated DNA samples (0%, 25%, 50%, 75%, 100% methylated; Zymo Research, Orange County, USA). Between chip variability was found negligible (Supplementary figure S6.2). All samples met the criteria tested with a fixed set of technical probes for bisulphite conversion, background signal, and hybridization efficiency, according to the Controls Dashboard from Illumina's GenomeStudio software [85]. Probe performance within each sample was investigated with the detection

P-value. It compares the signal generated at each CpG site to negative controls and can be interpreted as the probability of seeing a certain signal level without specific probe-target hybridization. Probes with detection P-value ≥ 0.05 in one of the samples were discarded for all samples before normalization with the Simple Scaling Normalization (SSN) method of the R-package *lumi* [237], which relies on signal intensity and methylation patterns (Supplementary figure S6.3).

In this study we focused on probes assessing CpG methylation on autosomal chromosomes, discarding the 11,650 probes on X- and Y- chromosomes. As annotated by the manufacturer, 86,769 contain a SNP (dbSNP 30) in their 50 bp target sequence or the CpG dinucleotide. Since it was shown on a earlier version of the platform that this could interfere with methylation measurements [238], we discarded measurements at these CpG site for further analyses. In all, we analyzed methylation measurements at 378,239 CpG sites in all six tissues for all six individuals.

Analyses

Comparing DNA methylation between the tissues

DNA methylation results from the Beadchip array platform can be described by two inter-exchangeable values, the ß-value and the M-value [239]. The ß-value represents DNA methylation on an interpretative scale of 0 to 1. The M-value represents it on a quantitatively more accurate logarithmic (log2) scale that resolves the skewed distribution in the upper (ß > 0.8; M > 2) and lower (ß < 0.2; M < -2) methylation levels [239]. For analyses and calculations the M-value was used for methylation levels of CpG sites, for presentation purposes the relevant results were transformed into the ß-value.

For genome wide DNA methylation patterns, principal component analysis (PCA) was used to investigate inter-individual variation vs. between-tissue (intra-individual) variation. At individual CpG sites, methylation differences between blood and the other tissues were tested with a two-sided t-test, and mean methylation, transformed into ß-value, and methylation variation, represented with the size of the standard deviation (SD) interval (mean ± 1*SD) in ß-value, were compared. Further, relative similarities between blood and the tissues were investigated using the Pearson

correlation coefficient (r). From here on in this study, tissue correlation, unless specified, refers to the correlation of DNA methylation in blood with that of another tissue, and tissue difference, unless specified, refers to differences in mean DNA methylation between blood and another tissue.

Smoothing tissue correlations and tissue differences

Previous research has demonstrated that methylation of nearby CpG sites is often correlated [195] and that this intra-class correlation can be used as a powerful tool to reveal subtle differences in DNA methylation that are consistently similar across several nearby CpG sites [135,233]. Exploiting this principle, we applied a smoothing algorithm with a sliding bandwidth of 2,000 bases centered around the cytosine residue of the CpG site under investigation. This created a more robust measure for tissue correlations and differences at individual CpG sites, partly compensating for the small amount of individuals in this study. After smoothing the absolute difference was used in combination with results from the two-sided t-test to describe tissue differences.
For tissue correlations we discriminated between CpG sites with fewer than 3 and those with 3 or more additional CpG sites within the bandwidth (Supplementary figure S6.4). The smoothed correlation at the latter CpG sites was considered indicative of a more robust local effect, representing the mean over 4 or more CpG sites. We focussed our exploration of tissue correlations on these CpG sites, which creates a bias for more CpG rich genomic areas, as CpG poor areas are less likely to meet these criteria. A smoothed correlation was considered strong when $r > 0.75$. Note that in this study all negative tissue correlations, even when strong ($r < -0.75$), were given the same weight as weak positive correlations, since, although mathematically equal to positive correlations, there is to our knowledge no experimental data in support of a biological interpretation of negative tissue correlation.

To scan for genomic regions enriched in tissue correlations we divided the genome in blocks with a length of 100,00 bp. For each block, we counted the number of probes with at least 3 additional CpG sites within the smoothing bandwidth (CpG rich area probes), the number of such probes with a strong tissue correlation, the proportion of correlated probes (for blocks ≥ 15 probes counted), the number of probes with strong tissue correlation between blood and 1 or 2 tissues,

between blood and 3 to 5 tissues, and between blood and all 6 tissues. We plotted this information against the human genome (assembly hg18/NCBI36) using the genome graph function of the UCSC genome browser [150] (website: http://genome.ucsc.edu/).

Identifying characteristics of tissue correlations

To investigate whether CpG sites with a strong correlation can be identified by characteristics of DNA methylation, we first explored the relation of tissue correlation with mean methylation, methylation variation, and tissue differences. For this means and SD interval sizes were categorized based on their observed distributions across all CpG sites, and tissue differences based on previously published observations on inter individual and between tissue differences [100,195,233]. Mean methylation categories: low (ß < 0.2), below medium (0.2 ≤ ß < 0.4), medium (0.4 ≤ ß < 0.6), above medium (0.6 ≤ ß < 0.8), and high (ß ≥ 0.8). Methylation variation categories: low (SD interval < 0.05), intermediate (0.05 ≤ SD interval < 0.15), and high (SD interval ≥ 0.15). Tissue difference categories: high (absolute difference ≥ 0.2 & $p_{t\text{-}test}$ ≤ 0.05), intermediate (0.05 ≤ absolute difference ≥ 0.2 & $p_{t\text{-}test}$ ≤ 0.05), low (absolute difference ≤ 0.05 & $p_{t\text{-}test}$ ≤ 0.05), and insignificant ($p_{t\text{-}test}$ > 0.05). The distributions across these categories were inspected for CpG sites from CpG poor areas, CpG sites from CpG rich areas with no strong tissue correlation and CpG sites from CpG rich areas with a strong correlation with at least one tissue. Note that a previous version of the platform already observed that CpG rich areas have different distributions of mean methylation and methylation variation compared with CpG poor areas [90].

To investigate whether CpG sites with a strong correlation can be identified by the function of their genomic area as annotated by the manufacturer [85] (Supplementary table S6.2, for full list of genomic locations) we inspected frequency distributions of these locations between strong correlations with 3 or more tissues, strong correlations with 1 or 2 tissues and weak correlations with all tissues in CpG rich areas, and also between strong and weak correlations in CpG poor areas.

To test for structural differences, CpG sites with the strongest correlations (r > 0.75 for at least 1 tissue correlation, and r > 0.25 for all tissue correlations; 6,539

CpG sites) were compared to CpG sites with the weakest correlations ($r < 0.25$ for all tissue correlations; 5,837 CpG sites), all located in CpG rich areas. A Chi^2 test was used to discern differences in the distributions of their genomic location (as annotated by the manufacturer). The Wilcoxon test of the epigraph web tool [240] (website: http://epigraph.mpi-inf.mpg.de/WebGRAPH/) was used to investigate differences in immediately adjacent DNA sequence (base composition and 2-mers, +/- 10 bp from C residue), location in transcription factor binding sites (TFBS) and repeats, and all available data on DNA structure, chromosome organization, and epigenome and chromatin structure (histone code in blood) [240].

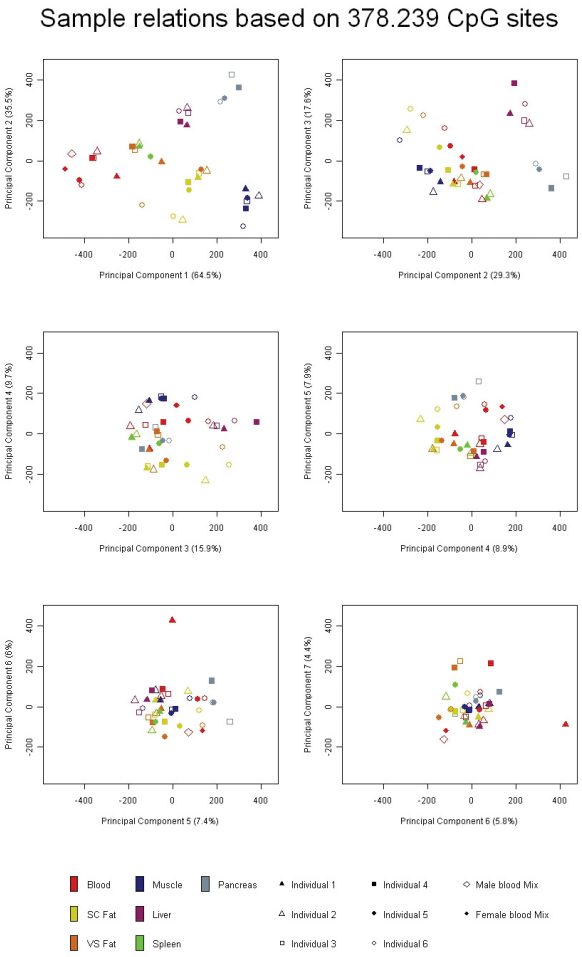


Figure 6.1: Scatter plots of principal components 1 to 7, derived from the Principal Component Analysis comparing the genome wide DNA methylation patterns of the tissue samples. In the top left corner the two lowest components (1st and 2nd) are plotted against each other, in the bottom right the two highest (6th and 7th). The lowest of two components is on the x-axis, the percentage giving on the axes is a measure of the amount variation represented by the component in this analyses. The symbols plotted are specific for samples of each individual, with open symbols for men, and a color specific to samples of each tissue. The distance between the samples is a measure of their similarity.

Results

PCA analysis on genome wide methylation patterns

A genome wide survey resulted in DNA methylation data across seven tissues of six individuals at 378,239 CpG sites distributed throughout the genome. Variation of genome wide patterns in DNA methylation was investigated by visualizing results from principal component analysis (PCA; Figure 6.1). Plotting the 1st vs. the 2nd principal components, covering almost all variation between the genome wide methylation patterns, showed tight clusters of the samples from the same tissue of each individuals (Figure 6.1, top left), even plotting the 6th vs. 7th principal component, covering roughly 10 % of variation, still showed no clusters of the samples from the various tissues of the same individual (Figure 6.1, bottom right). In this analysis the male and female blood mixes, representing healthy individuals, were observed to cluster together with the post mortem blood samples up to the 5th principal component, in the 6th and 7th components the blood mixes separated somewhat from the rest of the blood samples, but from the fourth component onward the analysis revealed no specific clusters of DNA samples (Figure 6.1).

Mean methylation and variation per tissue

The distribution of mean methylation across all 378,239 CpG sites investigated showed a similar bimodal shape in all tissues with peaks at low and high methylation and a valley at medium methylation (Figure 6.2). The distribution of methylation variation across all CpG sites also had a skewed shape similar for each tissue, with the SD interval peaking around 0.05 in ß-value and a long tail of higher variation, starting at SD interval around 0.2 in ß-value, and continuing to a maximum SD interval around 0.9 in ß-value (Figure 6.3).

Inspecting tissue differences at individual CpG sites revealed that almost 74 % of CpG sites showed a nominally significant ($p_{T\text{-test}} < 0.05$) difference in DNA methylation between blood and at least one of the other tissues (Table 6.2). Looking at the comparison of blood with each tissue, at the most 48 % of CpG sites, in the case of the spleen, and at the least 9 % of CpG sites, in the case of the pancreas, were differently methylated ($p_{T\text{-test}} < 0.05$). Most CpG sites with tissue differences had a different DNA methylation between blood and multiple tissues. For the spleen the biggest set of

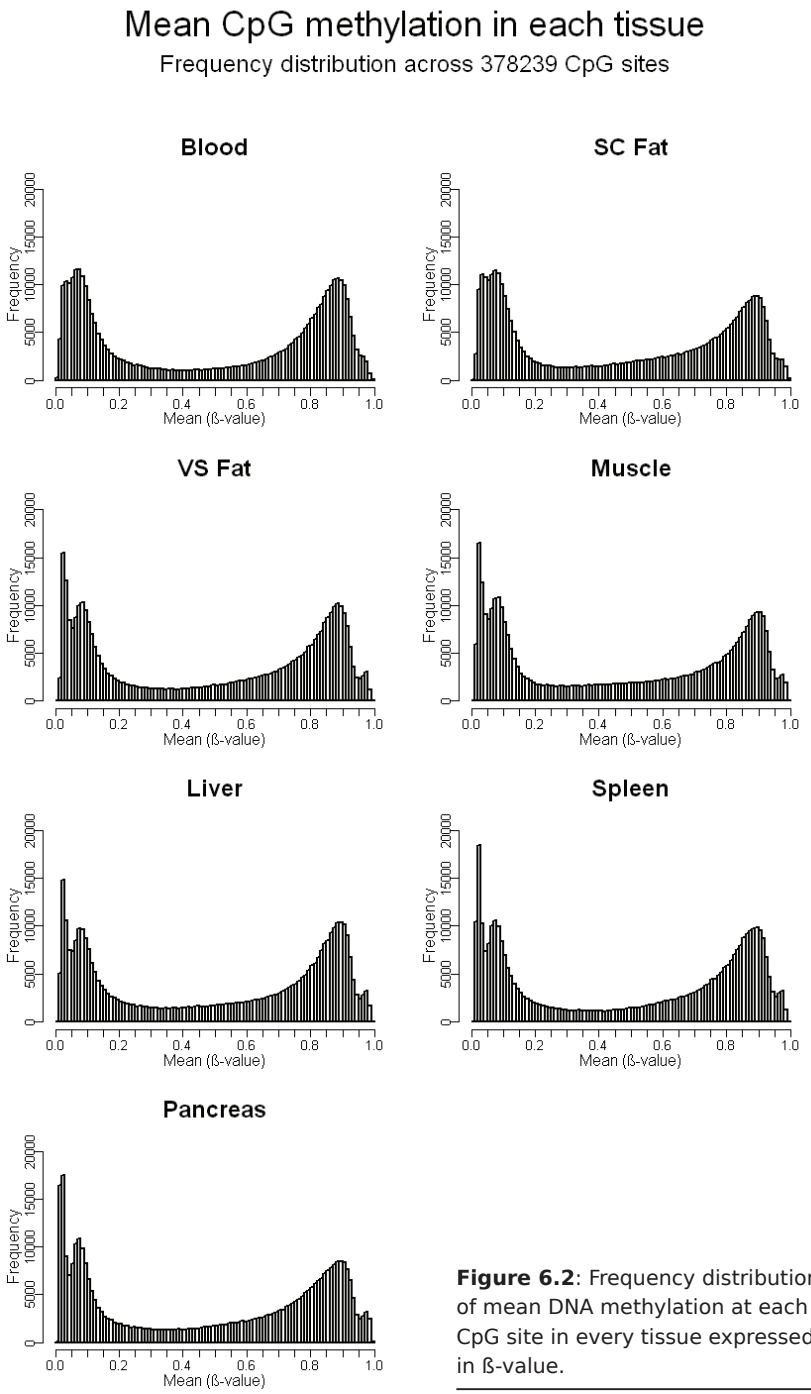


Figure 6.2: Frequency distributions of mean DNA methylation at each CpG site in every tissue expressed in ß-value.

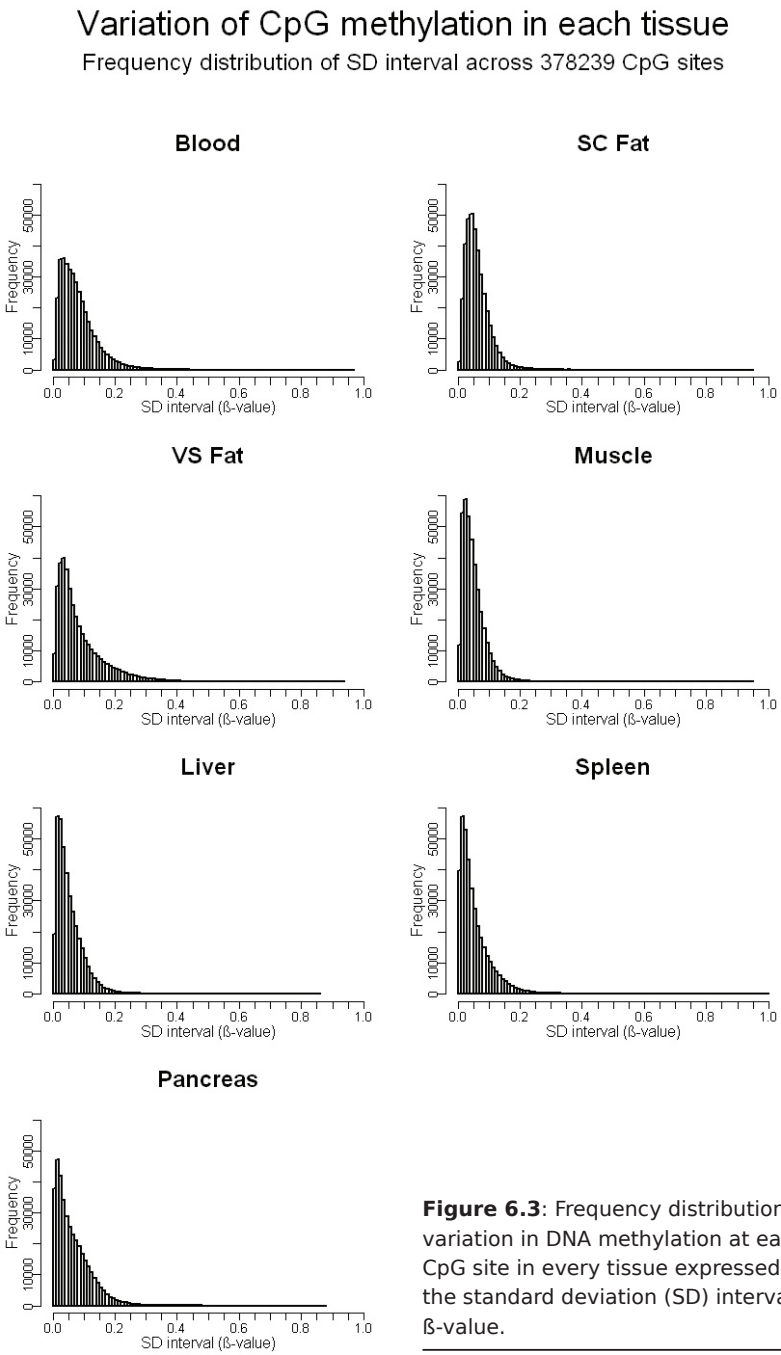


Figure 6.3: Frequency distributions of variation in DNA methylation at each CpG site in every tissue expressed as the standard deviation (SD) interval in ß-value.

Table 6.2: Significant methylation differences between blood and other tissues across 378,239 CpG sites

Amount of tissues[a]	Amount (percentage) of CpG sites with significant ($P_{T\text{-}test} < 0.05$) methylation differences													
	SC Fat		VS Fat		Muscle		Liver		Spleen		Pancreas		All tissues	
0	255426	(67.5 %)	295548	(78.1 %)	204087	(54.0 %)	247602	(65.5 %)	197537	(52.2 %)	344186	(91.0 %)	98618	(26.1 %)
1	5573	(1.5 %)	2533	(0.7 %)	18786	(5.0 %)	14357	(3.8 %)	44549	(11.8 %)	2619	(0.7 %)	88417	(23.4 %)
2	18005	(4.8 %)	6705	(1.8 %)	40505	(10.7 %)	24861	(6.6 %)	36786	(9.7 %)	3554	(0.9 %)	65208	(17.2 %)
3	30887	(8.2 %)	14427	(3.8 %)	42488	(11.2 %)	27055	(7.2 %)	33250	(8.8 %)	3723	(1.0 %)	50610	(13.4 %)
4	30201	(8.0 %)	20494	(5.4 %)	33546	(8.9 %)	26573	(7.0 %)	28530	(7.5 %)	4800	(1.3 %)	36036	(9.5 %)
5	24656	(6.5 %)	25041	(6.6 %)	25336	(6.7 %)	24300	(6.4 %)	24096	(6.4 %)	5866	(1.6 %)	25859	(6.8 %)
6	13491	(3.6 %)	13491	(3.6 %)	13491	(3.6 %)	13491	(3.6 %)	13491	(3.6 %)	13491	(3.6 %)	13491	(3.6 %)
Total[b]	122813	(32.5 %)	82691	(21.9 %)	174152	(46.0 %)	130637	(34.5 %)	180702	(47.8 %)	34053	(9.0 %)	279621	(73.9 %)

a: The number tissues with which these CpG sites have a different methylation compared with blood

b: The total amount of CpG sites that have a different DNA methylation between blood and this tissue

differently methylated CpG sites were uniquely so between blood and spleen, whereas for the pancreas the biggest set was differently methylated between blood and all six tissues (Table 6.2).

Correlations of CpG methylation between tissues

An alternative way to express tissue similarity, which takes into account differences in inter- and intra-individual (between-tissue) variation in DNA methylation across the tissues, is to compute correlation coefficients between blood and the other tissues. To inspect tissue correlations at a local genomic scale the correlation coefficients of each CpG site were smoothed over all measured CpG sites within the bandwidth of 2000 bp. For 92,916 CpG sites there were no additional CpG sites within the 2000 bp bandwidth (Supplementary figure S6.4). The other 285,323 CpG sites had from 1 up to 59 additional CpG sites within the bandwidth. The smoothed correlation was considered indicative of a local genomic effect when the bandwidth contained at least 4 CpG sites. The resulting smoothed correlation of such CpG sites thus represents a mean of measurements at 4 to 60 CpG sites. To illustrate this, the methylation values of blood against the other tissues were plotted for each of the 5 CpG sites within the bandwidth around probe cg2201694 (Figure 6.4).

Inspecting the smoothed correlations revealed that over 30 % of CpG sites had a strong correlation ($r > 0.75$) between blood and at least one other tissue. This was true for CpG rich (219,558 CpG sites) and CpG poor (158,681 CpG sites) areas (Table 6.3A). Looking at the distributions of tissue correlations for each tissue at the most 15.5 %, in the case of the spleen, and at the least 3.2 %, in the case of muscle, showed a strong correlation with blood. Remarkably, the endodermal tissues had more CpG sites strongly correlated with blood than the mesodermal tissues. Liver, the endodermal tissue with the lowest amount of CpG sites (15,444 sites) with a strong tissue correlation, still had almost 2,000 such CpG sites than SC fat, which was the mesodermal tissue with the highest amount CpG sites (13,501 sites) with a strong tissue correlation (Table 6.3A).

Looking in CpG rich areas at CpG sites with a strong correlation between blood and more than one tissue, there were 3 times more CpGs exclusively correlated between

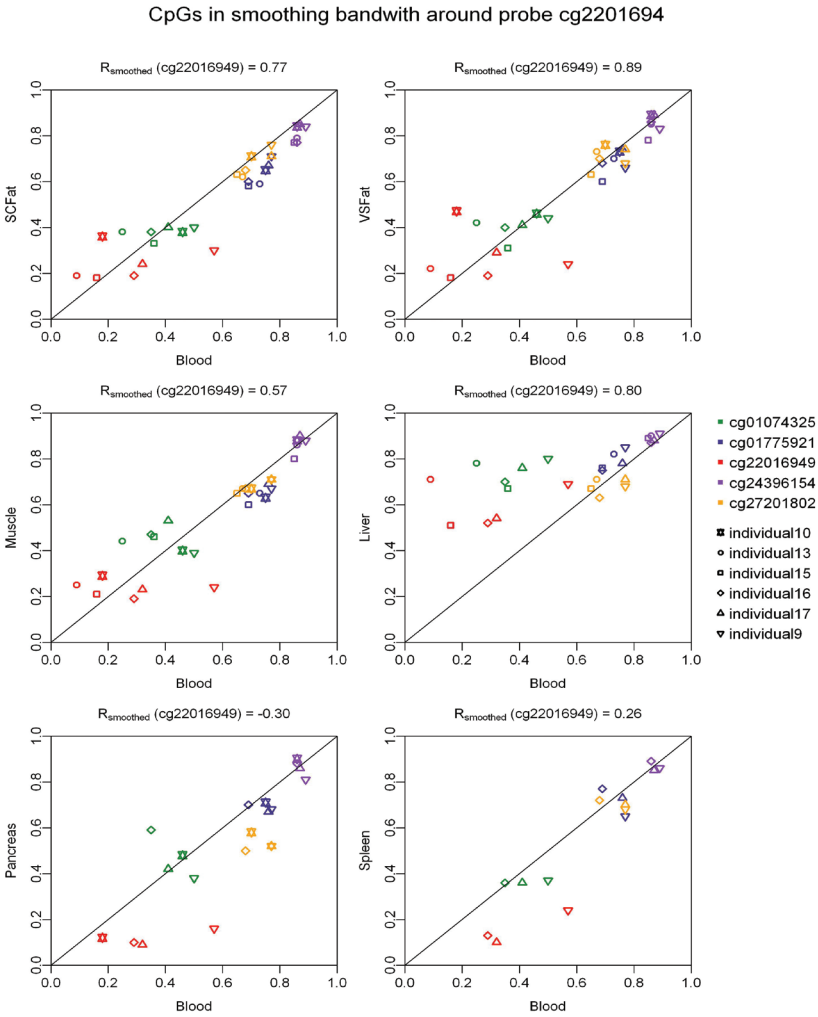


Figure 6.4: An example of smoothing tissue correlations over all CpG sites within a region of 2000 bp. The scatter plots show DNA methylation (in β-value) of blood (x-axes) against each other tissue (y-axes) per individual (plotted symbol) per CpG site (color) in the smoothing bandwidth around probe cg2201694. The smoothed tissue correlation of cg2201694 is given above each plot. The x = y line is given in black for reference.

Table 6.3A: Frequencies of CpG sites with strong and weak correlation between blood and the tissues in CpG rich and CpG poor areas

CpG density[a]	Correlated[b]	Amount (percentage) of CpG sites correlated between blood and each tissue[c]													
		SCFat		VSFat		Muscle		Liver		Spleen		Pancreas		All tissues	
CpG poor	Weak	148863	(93.81 %)	150683	(94.96 %)	153464	(96.71 %)	147441	(92.92 %)	134263	(84.61 %)	146658	(92.42 %)	109983	(69.31 %)
CpG poor	Strong	9818	(6.19 %)	7998	(5.04 %)	5217	(3.29 %)	11240	(7.08 %)	24418	(15.39 %)	12023	(7.58 %)	48698	(30.69 %)
CpG rich	Weak	206057	(93.85 %)	208456	(94.94 %)	212572	(96.82 %)	204114	(92.97 %)	185553	(84.51 %)	203143	(92.52 %)	152410	(69.42 %)
CpG rich	Strong	13501	(6.15 %)	11102	(5.06 %)	6986	(3.18 %)	15444	(7.03 %)	34005	(15.49 %)	16415	(7.48 %)	67148	(30.58 %)

Table 6. 3B: Frequencies of CpG sites from CpG rich areas with a strong correlation between blood and the tissues

Correlated[b]	Amount (percentage) of CpG sites correlated between blood and each tissue[d]													
	SCFat		VSFat		Muscle		Liver		Spleen		Pancreas		All tissues	
Strong with 1 tissue[e]	5447	(1.44 %)	3606	(0.95 %)	1986	(0.53 %)	6714	(1.78 %)	20266	(5.36 %)	8255	(2.18 %)	46274	(12.23 %)
Strong with 2 tissues[e]	4224	(1.12 %)	3781	(1.00 %)	2257	(0.60 %)	5117	(1.35 %)	9041	(2.39 %)	5144	(1.36 %)	14782	(3.91 %)
Strong with 3 tissues[e]	2092	(0.55 %)	2007	(0.53 %)	1228	(0.32 %)	2010	(0.53 %)	2985	(0.79 %)	1738	(0.46 %)	4020	(1.06 %)
Strong with 4 tissues[e]	875	(0.23 %)	847	(0.22 %)	679	(0.18 %)	776	(0.21 %)	945	(0.25 %)	586	(0.15 %)	1177	(0.31 %)
Strong with 5 tissues[e]	491	(0.13 %)	489	(0.13 %)	464	(0.12 %)	455	(0.12 %)	396	(0.10 %)	320	(0.08 %)	523	(0.14 %)
Strong with 6 tissues[e]	372	(0.10 %)	372	(0.10 %)	372	(0.10 %)	372	(0.10 %)	372	(0.10 %)	372	(0.10 %)	372	(0.10 %)
Strong mesodermal[f]	1355	(0.36 %)	1262	(0.33 %)	888	(0.23 %)							1682	(0.44 %)
Strong endodermal[f]							3994	(1.06 %)	5405	(1.43 %)	3956	(1.05 %)	6472	(1.71 %)

a: CpG rich area: 3 or more additional probes in smoothing bandwidth (2000 bp)
b: Weak correlation: $r < 0.75$; Strong correlation: $r \geq 0.75$;
c: Percentages are of all CpG sites within CpG poor areas (158,681 CpG sites) or within CpG rich areas (219,558 CpG sites)
d: Percentages are of all CpG sites examined (378,239 CpG sites)
e: CpG sites in CpG rich area with strong correlation between blood and number of tissues
f: CpG sites in CpG rich area with strong correlation between blood and more than one mesodermal / endodermal tissue

blood and the endodermal tissues compared with the mesodermal tissues (Table 6.3B). Although, a strong tissue correlation was observed between blood and one tissue exclusively for most CpG sites, more than 6,000 CpG sites showed a strong tissue correlation with three or more tissues, and 372 CpG sites with all tissues (Table 6.3B and, for each combination of tissues, Supplementary table S6.1). We then counted the amount of these tissue correlations per genomic fragment of 100,000 bp long (100K block) and scanned for genomic areas enriched for tissue correlation in a genome wide plot (Figure 6.5). We found that genomic areas with a higher amount of correlated probes (67,148 probes distributed over 11,080 100K blocks, plotted in blue) corresponded to areas with more measurements in general (219,558 probes distributed over 12,060 100K blocks, plotted in black). The proportion of correlated probes per 100K block (≥ 15 measurements; 162,014 probes distributed over 4,907 100K blocks, plotted in green) revealed no extended area with specific enrichment of tissue correlation, although there were 100K blocks with low (< 0.10) and with a high (> 0.50) proportion of correlated probes. In general one third of probes displayed a tissue correlation irrespective of measurement density, with a similar genome wide distribution of the amount of probes at which methylation of blood was correlated with one or two internal tissues (61,052 probes distributed over 11,077 100K blocks, plotted in purple), or with multiple tissues (5,674 probes distributed over 3,850 100K blocks, plotted in orange). At last, the distribution of the 372 probes correlated between blood and all internal tissues (distributed over 288 100K blocks, plotted in red) seems less related to measurement density.

Characteristics of correlated CpG sites

We inspected the distributions of CpG sites across the categories of mean methylation, methylation variation, and tissue differences to investigate whether strong tissue correlations could be marked by such characteristics of DNA methylation. We focused this inspection on CpG sites from CpG rich areas (219,558 sites), as their correlations represent a weighted average over multiple CpG sites. We observed no difference in these distributions between CpG sites with a strong tissue correlation compared with all CpG sites from CpG rich areas in any of the tissues (Tables 6.4 and 6.5).

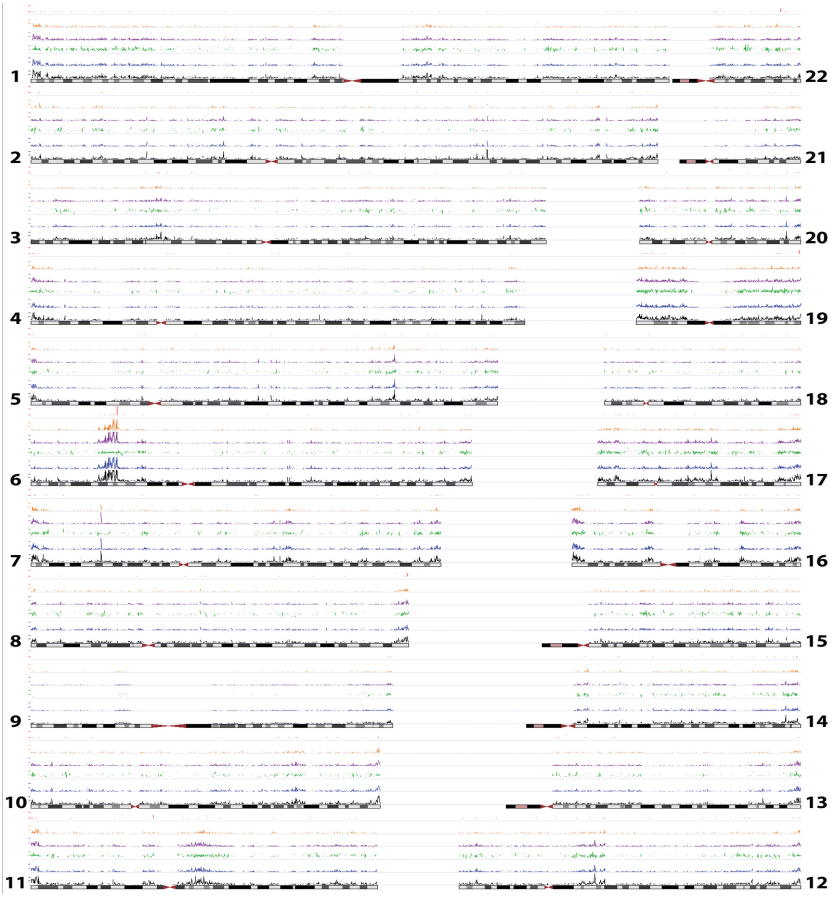


Figure 6.5: Plot of correlation of DNA methylation between blood and internal tissues (subcutaneous fat, visceral fat, skeletal muscle, liver, pancreas, and spleen) across the human genome (assembly: hg 18 / NCBI 36). Lines above each chromosome represent frequency information per unit of 100,000 bp long (100K block) of the number of probes with 3+ additional CpGs in the 2,000 bp smoothing bandwidth (black line), the number of probes with strong tissue correlation (blue line), the proportion of tissue correlated probes for blocks containing ≥ 15 probes (green line), the number of probes with tissue correlation between blood and 1 or 2 tissues (purple line), between blood and 3 to 5 tissues (orange line) and between blood and all 6 tissues (red line). The plot was made using the genome graph function of the UCSC genome browser [150] (website: http://genome.ucsc.edu/).

Table 6.4: Frequency distributions of mean DNA methylation and methylation variation for CpGs with strong and weak correlations

	CpG sites		Distribution (%) over mean methylation categories[a]					Distribution (%) over variation categories[b]		
Tissue	Category	n	0 – 0.2	0.2 – 0.4	0.4 – 0.6	0.6 – 0.8	0.8– 1	0 – 0.05	0.05 – 0.15	0.15 – 0.9
Blood	Strong CpG rich[c]	67148	53.9	8.9	5.6	10.9	20.6	44.5	46.4	9.2
	All CpG rich[d]	219558	56.0	8.9	5.5	10.2	19.4	46.0	44.9	9.1
	All CpG poor[e]	158681	6.9	5.0	7.7	25.8	54.6	19.6	66.9	13.5
SC Fat	Strong CpG rich[c]	13501	55.6	8.5	8.9	11.8	15.2	49.8	46.7	3.5
	All CpG rich[d]	219558	56.2	8.8	8.2	10.8	15.9	50.8	46.0	3.2
	All CpG poor[e]	158681	5.9	5.7	12.5	28.2	47.8	33.9	62.4	3.7
VS Fat	Strong CpG rich[c]	11102	52.6	9.2	7.9	12.7	17.7	46.4	39.4	14.2
	All CpG rich[d]	219558	54.3	9.4	7.6	11.0	17.6	50.1	36.9	13.1
	All CpG poor[e]	158681	5.1	4.6	10.2	28.6	51.5	28.0	50.7	21.4
Muscle	Strong CpG rich[c]	6986	56.3	9.2	8.6	10.0	15.8	68.0	30.1	1.9
	All CpG rich[d]	219558	56.2	9.0	8.0	10.3	16.5	68.3	29.9	1.8
	All CpG poor[e]	158681	6.4	7.7	12.7	24.7	48.6	46.7	50.4	2.9
Liver	Strong CpG rich[c]	15444	51.1	10.9	8.6	10.9	18.5	63.3	33.9	2.9
	All CpG rich[d]	219558	52.1	10.7	8.1	10.5	18.6	64.6	32.8	2.6
	All CpG poor[e]	158681	4.8	5.1	10.5	26.1	53.5	48.2	47.6	4.2
Spleen	Strong CpG rich[c]	34005	54.2	8.7	6.8	11.7	18.6	65.5	29.6	4.9
	All CpG rich[d]	219558	55.4	8.7	6.7	10.7	18.5	66.3	29.1	4.7
	All CpG poor[e]	158681	5.6	4.6	9.0	27.6	53.1	51.2	42.0	6.8
Pan-creas	Strong CpG rich[c]	16415	55.8	9.0	8.0	10.3	16.9	61.3	32.5	6.1
	All CpG rich[d]	219558	55.8	8.7	7.5	10.3	17.7	62.2	31.8	5.9
	All CpG poor[e]	158681	5.8	6.1	11.4	29.6	47.0	33.4	56.2	10.4

a: Lower to upper boundary of mean methylation for categories expressed in β-value

b: Lower to upper boundary of size of SD interval for categories expressed in β-value

c: CpG sites in CpG rich areas with a strong correlation ($r > 0.75$) between blood and at least one tissue, for blood this means all such CpG sites

d: All CpG sites in CpG rich areas, which had 3 or more additional CpG sites available for smoothing within the 2000 bp bandwidth

e: All CpG sites in CpG poor areas, which had 2 or less additional CpG sites available for smoothing within the 2000 bp bandwidth

Table 6.5: Frequency distributions of tissue differences in average DNA methylation for correlated and uncorrelated CpG sites

	CpG sites		Distribution (%) over methylation difference categories[a]			
Blood vs Tissue	**Category**	**n**	**Insignificant**	**0 – 0.05**	**0.05 – 0.20**	**0.20 – 1**
SC Fat	Strong CpG rich[b]	13501	72.5	16.2	9.1	2.2
	All CpG rich[c]	219558	73.0	18.0	7.5	1.5
	All CpG poor[d]	158681	59.9	26.9	10.9	2.2
VS Fat	Strong CpG rich[b]	11102	83.5	13.5	2.8	0.1
	All CpG rich[c]	219558	81.2	14.6	3.8	0.4
	All CpG poor[d]	158681	73.9	20.2	5.2	0.6
Muscle	Strong CpG rich[b]	6986	58.5	25.2	11.6	4.7
	All CpG rich[c]	219558	59.3	23.4	12.6	4.6
	All CpG poor[d]	158681	46.5	30.9	16.7	5.8
Liver	Strong CpG rich[b]	15444	70.5	14.6	10.9	4.0
	All CpG rich[c]	219558	70.0	19.6	8.3	2.2
	All CpG poor[d]	158681	59.2	26.5	11.4	2.8
Spleen	Strong CpG rich[b]	34005	51.9	36.6	10.7	0.9
	All CpG rich[c]	219558	53.1	37.8	8.5	0.6
	All CpG poor[d]	158681	51.0	39.8	8.6	0.6
Pancreas	Strong CpG rich[b]	16415	92.3	3.4	3.2	1.0
	All CpG rich[c]	219558	91.6	4.6	3.0	0.8
	All CpG poor[d]	158681	90.1	5.6	3.4	0.9

a: Lower to upper boundary of methylation difference for categories expressed in β-value, when $p_{T\text{-test}} < 0.05$

b: CpG sites in CpG rich areas with a strong correlation ($r > 0.75$) between blood and at least one tissue, for blood this means all such CpG sites

c: All CpG sites in CpG rich areas, which had 3 or more additional CpG sites available for smoothing within the 2000 bp bandwidth

d: All CpG sites in CpG poor areas, which had 2 or less additional CpG sites available for smoothing within the 2000 bp bandwidth

Different distributions were observed between CpG sites from CpG rich areas and CpG poor areas. Mean methylation in CpG rich areas displayed a similar bimodal distribution as for all 378,239 CpG sites (Supplementary figure S6.5), although the peak at low methylation (ß < 0.2) was substantially higher (> 50 % of CpG sites) than the peak at high methylation (ß ≥ 0.8; 15 % – 20 % of CpG sites). In contrast, mean methylation

in CpG poor areas displayed a skewed shape, with the tail at low methylation (< 7 % of CpG sites) and the peak at high methylation (> 50 % of CpG sites). These distributions were of similar shape in al tissues (Table 6.4). Methylation variation in all tissues was substantially higher in CpG poor areas compared with CpG rich areas (Table 6.4). Tissue differences also appeared more pronounced in CpG poor areas compared with CpG rich areas, except for spleen and pancreas (Table 6.5).

Finally, we investigated the possibility of identifying CpG sites with strong tissue correlations by the genomic function, structure, or genetic sequence of their surrounding area. We first inspected frequency distributions of functional areas and observed differences in the function of the location of CpG rich and CpG poor areas, but not of CpG sites with strong or weak tissue correlations, either in CpG rich or in CpG poor areas (Supplementary table S6.2). Comparing CpG sites with the strongest and the weakest tissue correlations (6539 and 5837 sites, respectively), we tested and found that neither repetitive DNA, nor RefSeq defined genetic components, nor CpG islands, nor most regulatory features were associated with strong tissue correlations (Table 6.6). We further tested and found that CpG sites with strong tissue correlations do not have different DNA structure or sequence compared with CpG sites with weak tissue correlations. Most differences between strongly and weakly correlated CpG sites were observed for epigenomic features (Table 6.6). Details of features with nominally significant ($p < 0.05$) differences between CpG sites with strong and weak tissue correlations are given in Supplementary table S6.3.

Discussion

In this study we surveyed DNA methylation at 380,000 CpG sites distributed throughout the genome across seven tissues of six individuals, with the aim to uncover loci at which DNA methylation in blood may mark that of inaccessible tissues [202]. Resembling results of previous studies [100–102], variation in genome wide DNA methylation patterns was largely determined by tissue differences. However, in line with results from recent studies [195,234], a substantial subset of individual CpG sites, showed similar

Table 6.6: Exploration of associations between tissue correlation and structural genomic and epigenomic features, or genomic location

Annotations by (test type)[a]	Genomic or epigenomic features investigated for association with tissue correlation[b]		
	Feature name	**Investigated**	**Associated**
epiGRAPH (wilcoxon)	Chromosome Organisation	27	3
	DNA Structure	63	0
	Epigenome and Chromatin Structure	107	23
	Conserved TFBS	260	8
	Repetitive DNA	162	0
Illumina (chi square)	Regulatory feature	8	2
	Relation to CpG island	5	0
	RefSeq group	6	0

a: Features and relative genomic locations using the epigraph web tool, or the annotations of the platform manufacturer

b: The number of features investigated vs the number of features with a nominally significant ($p < 0.05$) difference between CpG sites with strong and weak tissue correlations

average DNA methylation or a high correlation coefficient between blood and an internal tissue. High correlation coefficients were neither associated with levels of average DNA methylation in both tissues nor with methylation variation, nor with methylation differences between the tissues. Genetic sequence, structure and genomic location also appeared unassociated with tissue correlations. Hence, although we found loci for which blood can represent epigenetic marks of inaccessible tissue, we did not find a distinguishing feature by which such loci can be recognized. Still, cataloging the loci with methylation similarities between tissues, both those with comparable methylation levels and those with a high correlation of methylation variation, will indicate the loci at which an accessible tissue like blood

can be used as a marker tissue for disease related tissues that are inaccessible. Using a marker tissue for inaccessible tissues may help the design of epigenetic epidemiological studies and relate results to common diseases beyond the context of the tissue measured [98,99,202].

Despite the striking differences in genome-wide methylation patterns between the tissues, we also observed that average DNA methylation at individual CpG sites was more often similar than different between blood and each tissue, which was also reported on in a previous study [100]. Further, roughly a quarter of the CpG sites investigated showed a similar average DNA methylation level across multiple tissues (frequently hypo- or hyper-methylated). In line with previous studies [100–102], these results indicate the existence of many loci at which methylation levels will be similar between blood and an inaccessible tissue of interest. As tissue differences in methylation levels at CpG sites are often interpreted as indicative of a difference in epigenetic regulation of gene expression [101] and / or cell type differentiation [88,102]. Conversely, CpG sites with a similar methylation level between tissues might indicate a similar regulatory role of the epigenetic mechanism at the locus for both tissues. However, the aim of this study was to uncover loci at which the methylation status of blood marks that of inaccessible tissues for which the correlation of methylation variation between the tissues will be more relevant than the actual methylation level in either tissue.

We observed that a substantial subset of CpG sites showed a strong correlation between blood and another tissue. Previous studies have reported on inter tissue correlations at specific candidate loci [133,134,195], and here we present more of such data on a genome wide scale [234]. The majority of strong correlations were found between blood and one tissue exclusively, but a still sizeable minority showed a strong correlation across multiple tissues. A few CpG sites were strongly correlated across all tissues, however the most logical explanation for such seemingly soma-wide correlation of DNA methylation will be unknown genetic background variation, unless there is a biological reason to assume otherwise, such as genomic imprinting. Thus, for epigenetic case-control studies, not the similarity of average methylation, but a high correlation of methylation variation between the tissues will define loci at which blood can be a marker tissue for inaccessible tissues that are more relevant for the disease [119,195,234].

We focused our survey on correlated CpG sites that were smoothed over a minimum of four CpG sites, which were inevitably located in CpG rich areas. such CpG sites with a strong tissue correlation after smoothing are more likely to produce reliable assays in epigenetic epidemiology potentially marking the epigenetics of an internal tissue. However, the percentage of CpG sites with a strong correlation was similar between CpG rich and CpG poor areas. Thus, there is no reason to assume that by definition CpG poor areas will provide less suitable targets. This could be ascertained by validating the strong tissue correlations of CpG poor areas in a study with bigger sample sizes. Since DNA methylation changes progressively during cell differentiation [30] the differences between cell lineages that separated early during development are likely more pronounced. We therefore anticipated more tissue correlations between blood and the other mesodermal tissues compared with blood and the endodermal tissues [195]. However, we observed the contrary, which suggests that fewer DNA methylation changes occur before the separation of the germ layers than occur between the different cell lineages within a germ layer. On the other hand, there were fewer individuals sampled for each of the endodermal tissues, the influence of which cannot be estimated. Tissue correlations, either between blood and one tissue or across multiple tissues, were unrelated to mean methylation levels, methylation variation, or tissue differences at the CpG site, indicating that such characteristics of DNA methylation itself cannot be used to identify likely candidate CpG sites with strong correlation between marker tissue and disease related tissue. Correlation between DNA methylation in blood with that in other tissues appears to be complex and dependent on both locus and tissue, as previously suggested [195].

Design of epigenetic studies and interpretation of their results, would be greatly enhanced if genetic or genomic characteristics could be revealed by which the potential for tissue correlation of a locus can be recognized [119,195]. A good source for such characteristics may be related to the function of DNA methylation [24]. Repetitive DNA tends to be hyper-methylated in all tissues [39], which implies that DNA methylation at repetitive elements may be more discriminative between individuals [241]. In contrast, DNA methylation at CpG island shores was recently demonstrated

as especially discriminative between tissues [129]. It may be reasonable to expect CpG sites with a strong correlation between tissues to be located more frequently in repetitive elements and less frequently in CpG island shores. However, in neither of these features did we observe a different frequency of appearance for the most correlated CpG sites compared with the least correlated CpG sites. Similarly, we also did not observe any difference between CpG sites with strongest and weakest correlations with respect to their genetic sequence motifs, their genome structure or their location relative to genomic functional elements, for all of which the epigenetic state has been repeatedly related to cellular gene expression potential [21,89]. These results would suggest that correlation of DNA methylation between tissues may not be related to its functional aspects. On the other hand, our exploratory results indicate that a subset of epigenomic features in blood, such as histon modifications, particularly methylation at various lysine residues of Histon H3, and associations with RNA polymerase II, may be related to tissue correlations. This apparent contradiction cannot be resolved in our current study. To elucidate what causes a strong correlation of DNA methylation between tissues at certain loci, and by which features such loci can be identified will require substantial research efforts.

We compared CpG methylation of DNA from tissue samples collected during autopsy. DNA methylation is a covalent chemical bond that is stable over many years in stored DNA [195,197], and was shown to change little within the first few days after death [238], hence we set the maximum post mortem interval to 24 hours after death. In blood we found no indication that DNA methylation patterns were dramatically altered within these 24 hours after death, and there is no reason to assume this would be different for the other tissues. Thus the extrapolation of our observations to living individuals seems reasonable. Between chip variation of measurements was shown to be small and individuals whose DNA was analyzed on the same chip did not form clusters in the PCA. Data from probes with a known SNP in their 50 bp target sequence [238] was excluded from our analyses. Therefore technical artifacts seem an unlikely explanation for most of the observed tissue correlations, although the influence of (unknown) genetic variation in cis cannot be excluded [242,243]. To accommodate the

limitations of the small sample size of our study we employed a smoothing algorithm to create a median of the observed effects across the CpG sites within an area the size of 2000 bases. This will make the observation indicative of more robust locally carried effects, a method with demonstrated validity [135,233,244]. This will only work well at CpG sites if the recorded value can be internally validated by effects at sufficient nearby CpG sites. Thus, our observations at CpG sites in CpG poor areas lack the confidence granted by a sufficient sample size, and unfortunately have no other method available for increasing the robustness of the observed effect. This compelled us to adjust the comparison of strong and weak correlations to the CpG richer areas, which have different methylation characteristics (mean, variation and tissue differences), as observed in this study and in previous literature [90].

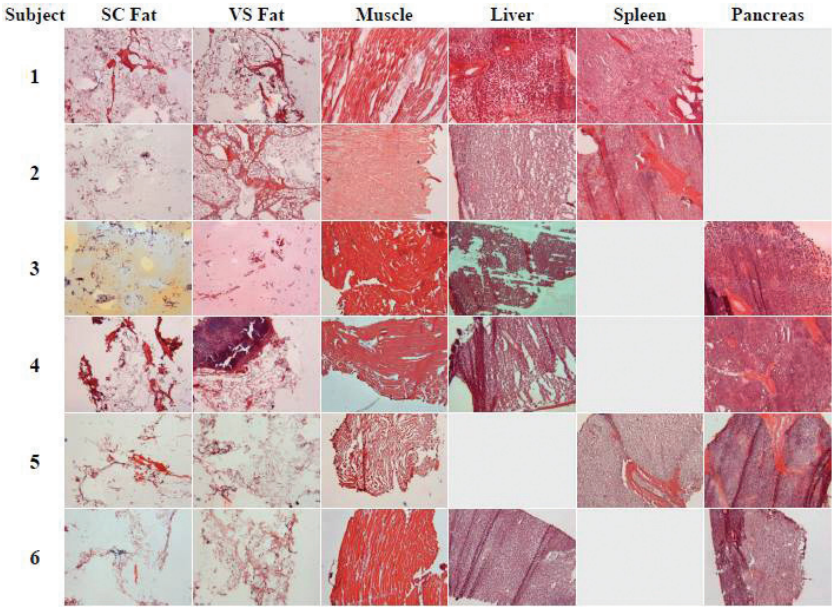
In this study we report that DNA methylation in blood may mark that of an internal tissues at a substantial subset of loci, which can be exploited for choosing assays and for interpreting results of epigenetic research. This capacity was established by strong correlation coefficients, which were complex in nature, depending on both locus and tissues. Strong correlations were unrelated to characteristics of DNA methylation, nor to the genetic sequence and genomic features surrounding the CpG site. Blood is the only tissue collected in most current human biobanks, due to its simple and cheap sampling techniques with low invasiveness. Other tissues that can also be collected in sufficient quantities for large populations with simple, inexpensive, and low invasive techniques are buccal swabs [245], epidermis [108], bladder lining from urine, and colonic mucosa from stool [195]. Such tissues may be used as marker tissues for the internal disease related tissues at different loci than are marked with blood. Although the effort to assess the usability maker tissues is substantial, epigenetic research on the common diseases will benefit greatly from the capacity to interrogate as many loci as possible, providing meaningful results through use of the right marker tissue (or more than one tissue for extra robustness) for that locus in combination with the disease related tissue.

Acknowledgement

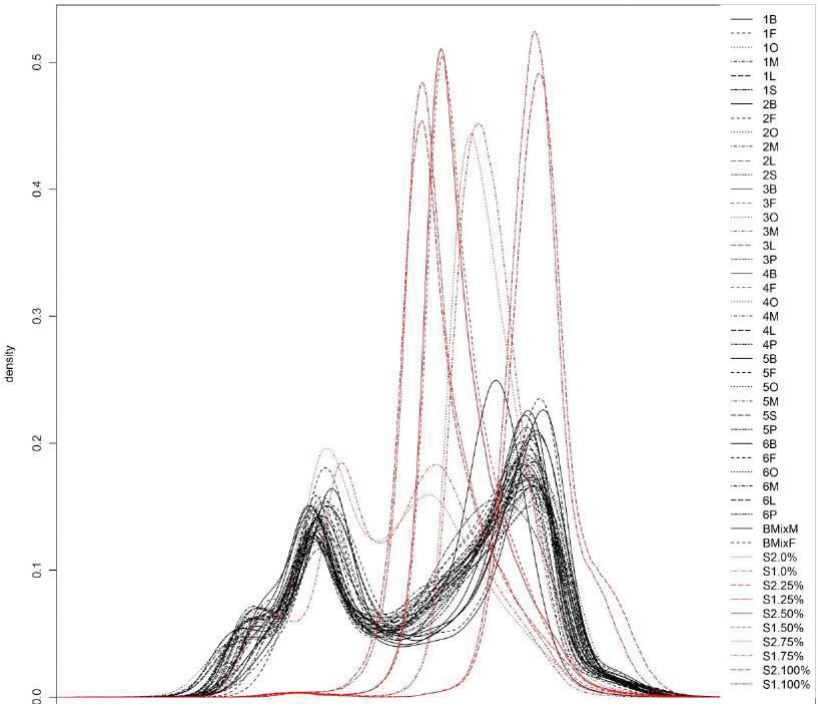
We would like to thank R. Guijt, T. Alberts, and E. Dreef for helping us obtain samples during the post mortem

investigations, and W. de Dijcker for his help in obtaining and processing the samples. The research leading to these results has received funding from the Netherlands Heart Foundation (2006B083), the European Union's Seventh Framework Program IDEAL (FP7/2007-2011) under grant agreement n° 259679, the Netherlands Consortium for Healthy Ageing (Grant 05060810) in the framework of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research, and Grant Number R01AG042190 from the National Institute On Aging.

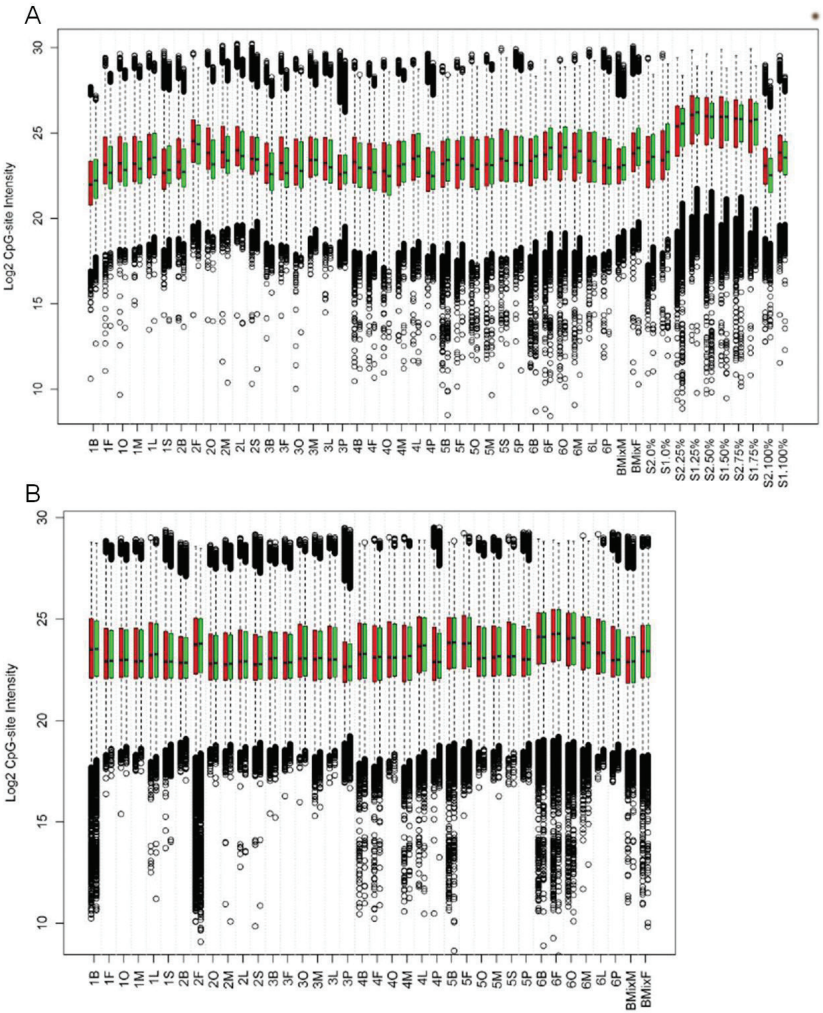
Supplementary Figures



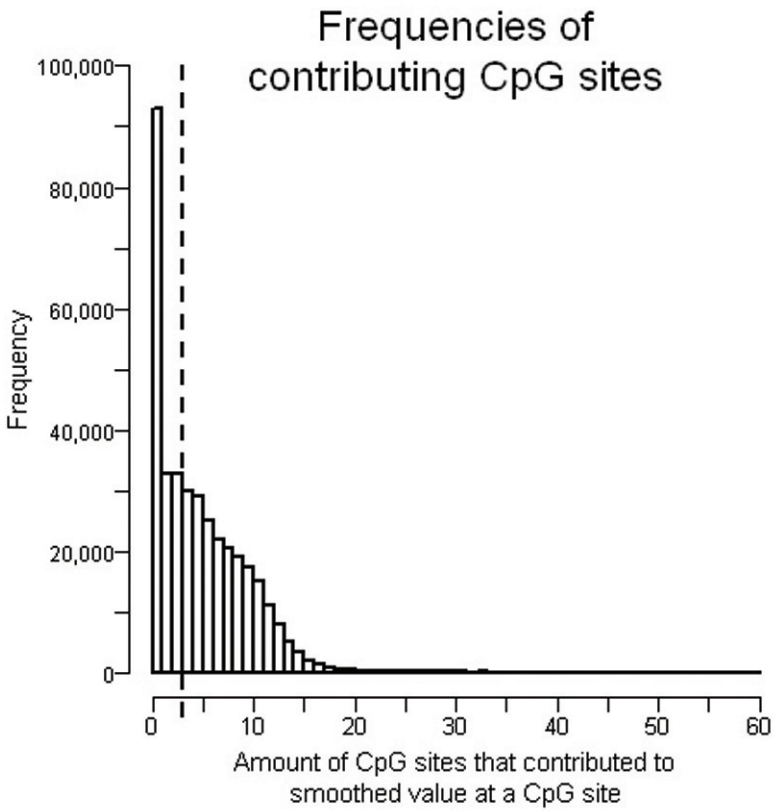
Supplementary Figure S6.1: Microscopy images of tissue coupes. Coupes are stained with HE, staining proteins and cytoplasm red and nuclei blue. SC Fat: subcutaneous fat; VS Fat: visceral fat. Except for visceral fat, the heterogeneity of the tissues appears similar between the individual



Supplementary Figure S6.2: Density plots of the raw M-values. Different methylation patterns are observed across all CpG sites between the tissue samples (black lines) and the methylation titer series (red lines). B:Blood, F:Fat, O:Visceral fat, M:Skeletal muscle, L:Liver, S:Spleen, P:Pancreas. 1-6: invididual 1 to 6, BMixM/F: Blood mixture Male/Female, S1/2.x%: Series1 or 2, x% methylated.



Supplemental Figure S6.3: Boxplots showing the intensity distribution (y-axis) of the two colour channels of the samples (x-axis) before (A) and after (B) normalization.
B:Blood, F:Fat, O:Visceral fat, M:Skeletal muscle, L:Liver, S:Spleen, P:Pancreas. 1-6: invididual 1 to 6, BMixM/F: Blood mixture Male/Female, S1/2.x%: Series1 or 2, x% methylated.



Supplementary figure S6.4: Distribution plot of the amount of CpG sites over which the correlation at a CpG site was averaged during smoothing. The vertical line represents the boundary between what are called in this study the CpG sites in CpG rich areas (3 or more additional CpG sites, thus correlation being a median over 4 CpGs) and the CpG sites in CpG poor areas.

Supplementary tables

Table S6.1: Frequency distribution of CpG sites in CpG rich areas with strong correlation (r ≥ 0.75) between blood and each combination of tissues

Amount of tissues	Amount of CpG sites	Combination of tissues					
		SCFat	VSFat	Muscle	Liver	Spleen	Pancreas
0	311091	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
1	5447	x					
1	3606		x				
1	1986			x			
1	6714				x		
1	20266					x	
1	8255						x
2	794	x	x				
2	420	x		x			
2	327		x	x			
2	853	x			x		
2	378		x		x		
2	303			x	x		
2	1505	x				x	
2	1648		x			x	
2	894			x		x	
2	2516				x	x	
2	652	x					x
2	634		x				x
2	313			x			x
2	1067				x		x
2	2478					x	x
3	141	x	x	x			
3	178	x	x		x		
3	112	x		x	x		
3	52		x	x	x		
3	461	x	x			x	
3	206	x		x		x	
3	182		x	x		x	
3	430	x			X	X	
3	343		x		X	X	

Amount of tissues	Amount of CpG sites	Combination of tissues					
		SCFat	VSFat	Muscle	Liver	Spleen	Pancreas
3	177			X	X	X	
3	93	X	X				x
3	69	x		x			x
3	83		X	X			x
3	161	x			x		x
3	100		x		x		x
3	46			X	X		x
3	241	x				X	X
3	374		x			X	X
3	160			x		X	X
3	411				X	X	X
4	74	X	X	X	X		
4	139	X	X	X		x	
4	195	X	X		X	X	
4	110	x		X	X	X	
4	73		X	X	X	X	
4	41	X	X	X			x
4	40	X	X		x		x
4	51	x		x	X		x
4	26		X	X	X		x
4	103	X	X			X	X
4	38	x		x		X	X
4	80		X	X		X	X
4	84	x			x	x	X
4	76		x		X	X	X
4	47			X	X	X	X
5	203	x	x	X	X	X	
5	127	X	X	X	X		x
5	68	X	X	X		X	X
5	59	X	X		X	X	X
5	34	x		X	X	X	X
5	32		x	X	X	X	X
6	372	X	X	X	X	X	X

The green background designates the mesodermal tissues and the blue background designates the endodermal tissues

Supplementary table S6.2: Distributions of genomic locations for CpG sites in CpG poor and rich areas with weak and strong tissue correlations

Genomic location[a]	Sites from CpG poor areas				Sites from CpG rich areas			
	Weak in 6T[b] (n = 109983)		Strong in 1T – 6T[b] (n = 48698)		Weak in 6T[b] (n = 152410)		Strong in 1T – 2T[b] (n = 61056)		Strong in 3T – 6T[b] (n = 6092)	
In relation to RefSeq gene										
none	41173	(37.4 %)	18257	(37.5 %)	21703	(14.2 %)	8834	(14.5 %)	900	(14.8 %)
1stExon	1154	(1.0 %)	490	(1.0 %)	11332	(7.4 %)	4665	(7.6 %)	423	(6.9 %)
3 UTR	6393	(5.8 %)	2944	(6.0 %)	2716	(1.8 %)	1057	(1.7 %)	93	(1.5 %)
5 UTR	7694	(7.0 %)	3331	(6.8 %)	15767	(10.3 %)	6367	(10.4 %)	604	(9.9 %)
Body	46985	(42.7 %)	20886	(42.9 %)	40377	(26.5 %)	16096	(26.4 %)	1639	(26.9 %)
TSS 1500	5453	(5.0 %)	2368	(4.9 %)	32681	(21.4 %)	13112	(21.5 %)	1360	(22.3 %)
TSS 200	1131	(1.0 %)	422	(0.9 %)	27834	(18.3 %)	10925	(17.9 %)	1073	(17.6 %)
In relation to regulatory feature and cell type specificity (CTS)										
none	93036	(84.6 %)	41855	(85.9 %)	72808	(47.8 %)	30403	(49.8 %)	3089	(50.7 %)
Gene Associated	389	(0.4 %)	161	(0.3 %)	318	(0.2 %)	144	(0.2 %)	12	(0.2 %)
Gene Associated (CTS)	720	(0.7 %)	276	(0.6 %)	376	(0.2 %)	166	(0.3 %)	15	(0.2 %)
Non Gene Associated	56	(0.1 %)	17	(0.0 %)	710	(0.5 %)	263	(0.4 %)	31	(0.5 %)
Non Gene Associated (CTS)	37	(0.0 %)	16	(0.0 %)	75	(0.0 %)	36	(0.1 %)	3	(0.0 %)
Promoter Associated	2622	(2.4 %)	998	(2.0 %)	50386	(33.1 %)	18854	(30.9 %)	1802	(29.6 %)
Promoter Associated (CTS)	816	(0.7 %)	276	(0.6 %)	2458	(1.6 %)	903	(1.5 %)	85	(1.4 %)
Unclassified	4539	(4.1 %)	1892	(3.9 %)	12712	(8.3 %)	5033	(8.2 %)	512	(8.4 %)
Unclassified (CTS)	7768	(7.1 %)	3207	(6.6 %)	12567	(8.2 %)	5254	(8.6 %)	543	(8.9 %)
In relation to CpG island										
Open sea	68053	(61.9 %)	30002	(61.6 %)	23978	(15.7 %)	9655	(15.8 %)	936	(15.4 %)
N Shelf	10553	(9.6 %)	4560	(9.4 %)	2638	(1.7 %)	1125	(1.8 %)	104	(1.7 %)
N Shore	8783	(8.0 %)	3937	(8.1 %)	25944	(17.0 %)	10286	(16.8 %)	1101	(18.1 %)
Island	6388	(5.8 %)	2967	(6.1 %)	77380	(50.8 %)	30901	(50.6 %)	3016	(49.5 %)
S Shelf	9724	(8.8 %)	4330	(8.9 %)	1979	(1.3 %)	846	(1.4 %)	83	(1.4 %)
S Shore	6482	(5.9 %)	2902	(6.0 %)	20491	(13.4 %)	8243	(13.5 %)	852	(14.0 %)

a: Information as supplied by manufacturer (Illumina, San Diego, USA)

b: 6T: in all six tissues; 1T – 6T: in one or more tissues; 1T – 2T: in one or two tissues; 3T – 6T: in 3 or more tissues

Supplementary table S6.3: Differences in structural features between CpG sites with the strongest and the weakest tissue correlations

Attribute Group Name	Attribute Name	$P^{a}_{wilcoxon}$	Weakest (n = 5837)		Strongest (n = 6539)	
			Mean	SD	Mean	SD
Chromosome Organisation	gieStain gpos50 overlap Average Size	.010	3911567	1769849	4131498	1760516
	gieStain gpos75 overlap Regions Count	.008	3.73	13.15	3.14	12.12
	gieStain gpos75 overlap Total Length	.008	74.70	262.92	62.70	242.44
Epigenome and Chromatin Structure	Overlap Regions Count	.003	54.19	77.77	58.57	80.98
	Overlap Total Length	.004	32.78	44.43	35.20	45.93
	tissue H1 overlap Regions Count	.003	27.15	41.54	29.55	43.39
	tissue H1 overlap Total Length	.003	27.15	41.54	29.55	43.39
	tissue imr90 overlap Regions Count	.018	27.03	41.81	29.03	43.50
	tissue imr90 overlap Total Length	.018	27.03	41.81	29.03	43.50
	cTissueHes hues8 oMeth ratio	.037	0.20	0.33	0.21	0.33
	chromMod CTCF overlap Regions Count	.007	22.71	104.89	20.88	112.90

Attribute Group Name	Attribute Name	$P^{a}_{wilcoxon}$	Weakest (n = 5837)		Strongest (n = 6539)	
			Mean	SD	Mean	SD
	chromMod CTCF overlap Total Length	.006	107.92	276.97	94.20	259.81
	chromMod H2A Z overlap Regions Count	.002	72.97	171.05	65.12	156.06
	chromMod H2A Z overlap Total Length	.003	298.63	417.45	277.08	408.61
	chromMod H3K-27me1 overlap Average Size	.031	24.33	0.45	24.28	0.45
	chromMod H3K-36me1 overlap Regions Count	.024	10.41	24.44	9.28	22.64
	chromMod H3K-36me1 overlap Total Length	.027	107.28	266.07	97.13	253.34
	chromMod H3K4me3 overlap Regions Count	$5.4*10^{-05}$	379.11	775.23	336.92	714.18
	chromMod H3K4me3 overlap Total Length	$9.9*10^{-05}$	547.73	462.54	516.12	463.05
	chromMod H3K-79me1 overlap Regions Count	.003	13.88	27.62	12.48	26.52
	chromMod H3K-79me1 overlap Total Length	.005	138.20	292.36	126.51	282.59
	chromMod PolII overlap Regions Count	$1.1*10^{-06}$	51.54	140.81	44.68	136.63
	chromMod PolII overlap Total Length	$8.1*10^{-07}$	252.51	394.28	219.14	375.02

a: Unadjusted p-value from Wilcoxon (epigraph webtool) or Chi Square test, only attributes with nominally significant ($p < 0.05$) differences are given

Supplementary table S3 (continued)

Attribute Group Name	Attribute Name	$P^a_{wilcoxon}$	Weakest (n = 5837) Mean	SD	Strongest (n = 6539) Mean	SD
Epigenome and Chromatin Structure	Overlap Average Size	.006	0.48	0.01	0.48	0.01
	Overlap Regions Count	$9.5*10^{-06}$	823.96	1041.14	761.10	981.93
	Overlap Total Length	.017	922.32	216.52	912.44	233.22
Conserved TFBS	cNameV ahrarnt 02 oZscore	.017	1.98	0.28	2.33	0.49
	cNameV creb 02 oZscore	.004	1.76	0.13	2.36	0.42
	cNameV elk1 01 oZscore	.030	2.58	0.59	2.13	0.43
	cNameV hen1 02 oZscore	.049	2.06	0.25	1.91	0.27
	cNameV myognf1 01 oZscore	.032	2.09	0.35	1.89	0.26
	cNameV ncx 01 oZscore	.032	2.62	0.54	1.99	0.31
	cNameV stat1 01 oZscore	.010	2.35	0.32	1.95	0.26
	cNameV tcf11mafg 01 oZscore	.017	2.27	0.29	1.97	0.26
Distributions of Illumina annotated genomic location		$P^a_{chi\ square}$	**Observed**	**Expected**	**Observed**	**Expected**
Location to Regulatory Feature	Promoter Associated	$6.1*10^{-04}$	1953	1846.0	1961	2068.0
	Unclassified Cell type specific	.026	462	498.1	594	557.9

a: Unadjusted p-value from Wilcoxon (epigraph webtool) or Chi Square test, only attributes with nominally significant ($p < 0.05$) differences are given

Chapter 7
Thesis Summary and discussion of results

Table 7.1: Summary of thesis results at candidate loci

	Characteristics of CpG methylation						Prenatal famine
	Level (%)		Cell population		Correlation		Period
Locus[a]	Mean	SD	Association[b]	Adjustable[c]	Longitudinal[d]	Blood vs. buccal[e]	P[f]
NR3C1	< 20	< 3.0	-				-
TNF	< 20	< 3.0	+	+			-
APOC1	< 20	3.0 – 6.0	+	+	+	+	-
IL10	20 – 40	> 6.0	+	-	-	-	↑
LEP	20 – 40	3.0 – 6.0	+	+	+	+	↑
KCNQ1OT1	20 – 40	< 3.0	-		-	-	-
ABCA1[m]	20 – 40	> 6.0	+	+			↑
GNASAB	20 – 40	3.0 – 6.0	-				-
GRB10	40 – 60	3.0 – 6.0	-				-
GNASAS	40 – 60	3.0 – 6.0	-				↑
MEG3	40 – 60	< 3.0	-				-
IGF2	40 – 60	3.0 – 6.0	-		+	-	↓[n]
CRH	60 – 80	> 6.0	+	+	+	+	-
IGF2R	60 – 80	> 6.0	-		+	+	-
INS	> 80	< 3.0	-		-	-	↓
FTO	> 80	< 3.0	-				-

a: Loci are ordered on mean level of methylation from lowest to highest

b: Significant influence of cell population heterogeneity on methylation variation: + = $p < 0.05$; - $p > 0.05$ (uncorrected for multiple testing)

c: The amount of influence of cellular heterogeneity is sufficiently small to allow statistical adjustment for it

d: Correlation of methylation variation between longitudinal blood samples (follow-up = 20 years): - = $\rho < 0.75$; + = $\rho > 0.75$

e: Correlation of methylation variation between blood and buccal cells: - = $\rho < 0.75$; + = $\rho > 0.75$

f: DNA methylation is associated with famine in the period around conception

g: DNA methylation is associated with famine in the last trimester of gestation (and early post natal life)

			Accumulated ageing related changes				Risk for CHD[l]	
	Specificity		Variation increase[j]		Environmental[k]			
L[g]	Sex[h]	Time[i]	Inter-ind-ivid-ual	Within-MZ-pair	Unique	Familial	Associa-tion	Sex specific
-		-						
-	-	+					-	-
-	M	-	++	++	+	-	-	-
-		-	-	+	+	-		
	-		+	++			-	-
↓	F	+	+	-	+	-	↑	F
-[n]	-[n]	+[n]	++	++	+	-	-	-
-		-	++	++	-	+		
-		-						
-	M	+	++	++			↑	F

h: The association with famine around conception is restricted to men (INS and LEP) or more pronounced in women (GNASAS)

i: The association with famine is restricted to only one period, or for GNASAS is in opposite direction between both periods

j: Significant (+) or substantial (++) increase of methylation variation (inter-individual and within pair discordance) calendar age

k: Attributability of the increase in methylation variation to the unique (individual) or the familial (shared) environment

l: Significant association of DNA methylation at the locus with risk of MI during 3 year follow-up is restricted to the women of the study

m:Information is on the methylated CpG sites at the 5'end of the assay (Chapter2)

n: These observations [14] are not included as a chapter of this thesis but were the study directly preceding Chapter 3

Epigenetic mechanisms regulate cellular gene expression potential without changing the genetic code [29]. Like the genetic sequence, epigenetic marks are faithfully transmitted during mitosis and are generally stable in differentiated cells, but in contrast with the static genome, the epigenome retains the capacity for dynamic changes in each individual cell. Epigenetic variation is therefore a topic of interest for research on ageing and its related common diseases [47,50]. In this thesis we focus on DNA methylation, which is the most studied layer of epigenetic information [28], and is correlated to the other epigenetic layers [21,25]. We used a combination of successive studies to investigate aspects of variation in DNA methylation, various sources generating such variation and its relation with risk for myocardial infarction (MI) at candidate loci for cardiovascular and metabolic diseases.

Thesis results

In **chapter two** we investigated the possibilities of epigenetic epidemiological research using samples stored in existing bio-repositories, most of which were originally designed for genetic research and usually contain DNA from blood drawn at baseline without longitudinal sampling. With reference epigenome browsers still under construction [88,89], we first established a reference of inter-individual variation in the methylation status of 104 CpG sites distributed over 16 loci (Table 7.1). These loci were selected because they map to candidate genes for cardiovascular and metabolic diseases and their epigenetic characteristics represent various potential targets for epigenetic regulation of gene expression (Chapter 1). We established that CpG methylation in the cell population of a tissue sample is a quantitative trait at these loci. We further found that, although both average methylation level and the amount of variation could vary substantially between adjacent CpG sites, the methylation of nearby CpG sites was generally correlated, which may have some practical analogy to the linkage disequilibrium blocks of genetic variation [159] and fits the common view that the regulatory effect of DNA methylation usually relies on the status of multiple CpG sites within a genomic region [94]. A recent whole methylome study confirmed the existence of such correlations and showed that they extent roughly a hundred base pairs to

either side [83]. We then investigated the concern that inter-individual variation in DNA methylation might be a reflection of leukocyte population heterogeneity [103]. We found that variation in DNA methylation at 10 loci was not influenced by leukocyte population heterogeneity, this influence was marginal at 5 loci, and it was substantial only at the *IL10* locus (Table 7.1). We described a statistical method to account for minor influences of cell type heterogeneity on DNA methylation variation. Recent studies have observed similar proportions of unaffected, marginally affected and strongly affected loci in a study on different candidate loci [246] and on a large set of individual CpG sites distributed throughout the genome [77].

Next we investigated the stability of variation in DNA methylation over time in longitudinal DNA samples of both blood (10-20 year time span) and buccal cells (2 – 8 year time span) in 34 individuals (Table 7.1). Although DNA methylation was not perfectly maintained at any locus, 5 of 8 loci assessed displayed a strong correlation ($\rho \geq 0.75$) between the time points in both tissues, indicating that prospective studies on DNA methylation are possible. Such locus specific temporal stability was also found in another longitudinal study [197]. DNA methylation levels of imprinted loci are commonly thought to be similar between tissues, since they originate in early development [69,106]. We finally investigated whether the variation in DNA methylation of blood may correspond to that of another tissue. We found that DNA methylation at 4 of 8 loci displayed a strong correlation ($\rho \geq 0.75$) between the recent blood and buccal cell samples, irrespective of imprinting status. A recent genome wide study reported indicated that DNA methylation patterns in blood could be a surrogate marker for cerebellum at some loci [234]. In all, depending on the locus of interest and with careful study design, some epigenetic research questions can be answered using DNA from existing biobanks.

In **chapter three** we investigated whether in utero exposure to the Dutch Famine of 1944-1945 is associated with persistent differences in epigenetic information at the candidate loci described in chapter 2, assessing the generality of our primary findings at the *IGF2* locus [14]. We compared DNA methylation between middle-aged individuals

who were born or conceived during the Dutch Famine and their same sex older or younger siblings. We found that DNA methylation was persistently altered in prenatally exposed individuals at 6 of the 15 loci assessed (Table 7.1). The direction of these exposure related changes was specific for the loci, with 5 showing higher, and 1 lower DNA methylation. Furthermore, the exposure related change in DNA methylation was sex-specific at 3 loci, and specific to the timing of the exposure at 5 loci. At 4 of these loci only famine exposure in the period around conception was associated with altered DNA methylation, whereas at the *GNASAS* locus famine exposure in both the period around conception and in the final three months of gestation resulted in altered DNA methylation in opposite directions. The combination of epigenetic changes specific to the locus, the type and timing of the exposure and the sex of the exposed can unlikely be explained solely by epigenetic damage due to a deficiency in methyl donors. Further research may reveal whether and which other mechanisms underlie these observation, including the proposed possibility of an adaptive response being recorded in the epigenome [178].

More prenatal conditions have been reported to produce persistent epigenetic changes, including maternal micronutrient deficiency [225], folate supplementation [120,219], maternal smoking [67,226], and placental morphology [217]. Such observations in humans resemble results from intervention studies in animal models [13,63,184]. In animal studies the epigenetic changes were also associated with phenotypic changes related to disease development, forwarding epigenetic alterations as a plausible molecular mechanism for the developmental origins hypothesis of adult disease, which poses that adverse conditions during development and early life increase the risk for age related common diseases like CHD [165]. However, since the Dutch Famine related epigenetic differences were observed in blood, it remains unclear whether and how they are involved in the phenotypic observations related to the Dutch Famine that indicate a predisposition to metabolic and cardiovascular diseases [167,168]. In any case, the persistence and specificity of epigenetic information found associated with early life exposures makes such marks attractive potential reporters for developmental conditions, for which the genetic sequence itself cannot provide any clue.

Further, epigenetic marks found associated with conditions around conception and early embryogenesis might even correlate across the tissues that developed from the same progenitor stem cells affected at the time of the exposure, thus potentially marking the epigenetic state of inaccessible tissues [69,94].

In **chapter four** we investigated the susceptibility of epigenetic information to changes during the whole adult life span, having observed in Chapter 2 that every locus assessed displayed some degree of epigenetic changeability over 20 years. This chapter aims to extend findings of previous studies investigating this question with cross-sectional and longitudinal designs on smaller sample sizes with narrower age ranges [130,197,247]. In 230 monozygotic twin pairs (MZ twins) ranging from young to elderly adults (18 – 88 years old) we compared the amount of inter-individual variation and within-pair discordance in both global DNA methylation and DNA methylation at 7 of the loci, assessed in chapter 2. Increases in these measures of epigenetic variation in the population were used as a proxy for combined stochastic and environmentally driven epigenetic changes in individuals and were both found to be largely independent from leukocyte population heterogeneity. Although at all ages global DNA methylation showed little variation, its variation increased significantly with age, and results at most candidate loci further indicated that epigenetic changes accumulate gradually with increasing calendar age. Recent genome-wide studies also report associations of DNA methylation with age [58,248]. The rate of accumulation appeared specific to the locus with some loci (e.g. *IGF2*, *INS*, *LEP*) displaying a substantial accumulation of (random) epigenetic changes throughout adult life, while inter-individual variation at *KCNQ1OT1* and within-pair discordance at *GNASAS* remained stable (Table 7.1).

At the *CRH* locus mean methylation levels were similar at all ages while increases in variation were substantial, potentially indicating that changing methylation marks may yield another trait in epigenetic research. Epigenetic instability is proposed to be both a stochastic initiator of disease development [50] and an important biological mediator of environmentally driven disease risk [68]. The age related increase in DNA methylation variation fits both

slightly contrasting views. Stochastic DNA methylation changes without an environmental component are likely related to imperfect epigenetic maintenance processes during mitosis [51,205]. However, the directed DNA methylation changes that result from environmental exposures related to lifestyle and living conditions [72,77] may likely accumulate as changes in (partly) opposite directions, since conditions change across a lifetime. Both processes will manifest as random epigenetic changes with age, which is commonly called epigenetic drift. We observed that most epigenetic variation and most of its age-related increase was generally attributable to individual factors (Table 7.1), which fits both causes for epigenetic instability. Remarkably, although longitudinal examination showed that the increase in methylation variation was continued into old age, there were still old twin pairs with virtually identical DNA methylation signatures, indicating that either the epigenome of some individuals is less changeable, or that these individuals are better capable of correcting epigenetic changes. The eventual build-up of dysfunctional epigenetic changes in each cell is thought to result in an age related increase in the proportion of cells in a tissue that are epigenetically dysregulated [50,205], a state that has been forwarded as a possible molecular mechanism in the process of ageing [58,59]. Our observations at old age may thus forward epigenetic marks in the elderly as a potential indicator of biological age [249], rather than calendar age [250]. Future research into the causes and phenotypic consequences of epigenetic instability may be able to shed light on such questions.

In this chapter we examined 5 of the 6 loci at which DNA methylation was found associated with prenatal famine [14]. We found that 4 of these loci showed considerable susceptibility to epigenetic drift, which would seem to contradict the observed persistence of the epigenetic changes induced by the Dutch Famine (Table 7.1). However, observations at the *IGF2* locus showed that the differences induced by the Dutch Famine were not similar for every exposed and non-exposed sibling pair, with the persistence of the difference manifested at the group level [14]. We observed in this chapter that there were still old twin pairs with similar methylation levels, and, although the loci show substantial increases in variation with age, changes

in the population averages seem proportionately smaller across a lifespan. A plausible interpretation of this would be that epigenetic differences related to the (prenatal) environment might be most pronounced in the population just after the event, and then slowly fade as epigenetic variation around the average increases with age through accumulated (random) epigenetic changes in each individual. An alternative explanation may view the Dutch Famine associated differences in DNA methylation as a persistently accelerated DNA methylation age in the prenatally exposed group, with the age related increase in epigenetic variation at these loci relating to an increased variation in DNA methylation age [251]. Both interpretations fit the developmental component and the stochastic variation of ageing [59,178].

In **chapter five** we investigated whether, in line with the developmental origins hypothesis [165], DNA methylation was related to risk for CHD at the 6 candidate loci, at which it was associated with prenatal famine [14] (Chapter 3). Within the placebo group of the PROSPER trial on Pravastatin and the risk of CHD [118,218], we compared all individuals who were event free at baseline and developed myocardial infarction (MI) during 3 years follow-up (n = 122) with a similar sized control group. Due to our results from the Dutch Famine study, we tested for and found a significant sex interaction for *INS* and *GNASAS*. Higher DNA methylation at these loci was associated with an increased risk for MI among women (Table 7.1) with the association strongest for individuals displaying hypermethylation at both loci. We observed no epigenetic associations among men, which could be due to an aspect of study design such as inclusion age [230] and / or may have a biological reason. This study was one of the first to show a locus specific epigenetic association with CHD, but more are emerging. A case-control study [252] on coronary artery disease (CAD) found an association with methylation of a candidate gene in the 9p21 locus, implicated in CHD by genome wide association studies [253]. A smaller retrospective study on familial hypercholesterolemia patients [254] reported an association between CAD and methylation at the exact same CpG sites of *ABCA1* as measured in this thesis. Our prospective study on MI risk found no association at *ABCA1*, the simplest explanation for which

would be differences in the study population (event free vs. hypercholesterolemic) and the study design (prospective vs. retrospective). The associations we found were independent of the well-established traditional risk factors of CHD (i.e. lipid profile, hypertension, Type 2 diabetes), suggesting that DNA methylation at specific loci may be an epigenetic marker for disease risk. This view has been substantiated by two recent longitudinal studies reporting associations in blood between persistent DNA methylation at candidate loci and obesity or type 2 diabetes [48,49]. With our investigation limited to blood, our results do not directly implicate an epigenetic mechanism in disease development.

Do results in this chapter reflect a developmental component of CHD? The original developmental origins hypothesis (Figure 7.1 A) was formulated on epidemiological studies in which various proxies for poor conditions in early life were associated with an increased incidence and severity of CHD in middle age [8,165]. In line with this hypothesis, epidemiological studies on both Dutch Famine cohorts [167,168] indicate that prenatally exposed individuals have a higher incidence of its phenotypic pre-stages (Figure 7.1 C). Results from a wealth of intervention studies on animal models [66] show that adverse early life conditions result in epigenetic changes specific to the insult and the locus, and that these associations coincide with proxies for the pre-stages of CHD (Figure 7.1 B). Both *INS* and *GNASAS* methylation is important in regulating fetal growth [180,201]. Chapter 3 showed that persistently different methylation at *INS* and *GNASAS* was associated with prenatal exposure to the Dutch Famine (Figure 7.1 C). In this chapter we observe that higher methylation at both loci is associated with risk for CHD independently of the traditional CHD risk factors (Figure 7.1 D). The associations at *GNASAS* are a complete match between prenatal (peri-conception) susceptibility and risk for CHD, both in terms of direction of the difference as in terms of sex-specificity (Table 7.1). Chapter 4 shows that methylation at *GNASAS* is mostly inert to epigenetic drift, making an unknown developmental component the most likely explanation for its association with CHD. The associations at *INS* are in opposite direction and in the other sex, prenatally exposed men have lower DNA methylation and women at risk for CHD have higher DNA methylation. Further, *INS* methylation is susceptible to changes, stochastic

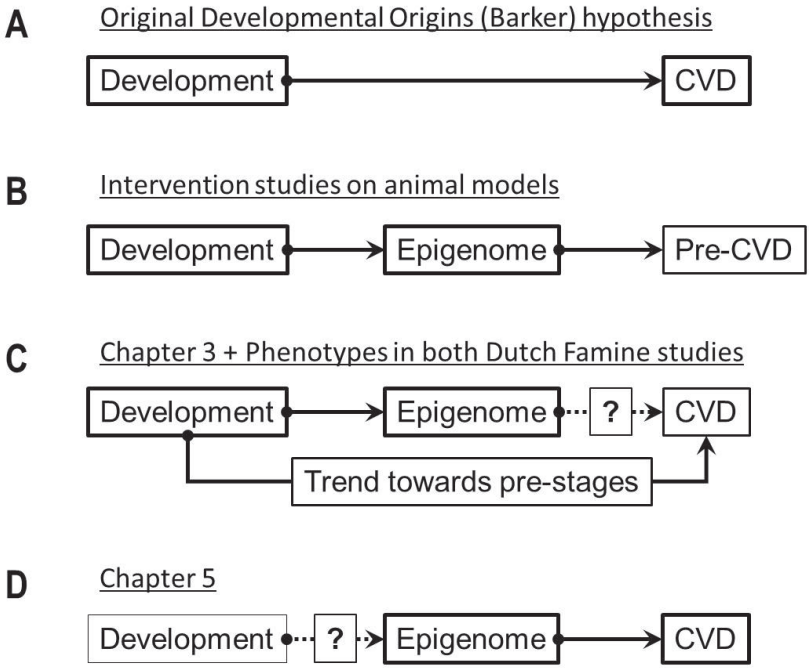


Figure 7.1: **Schematic summary of results related to DOHaD. A**) The original developmental origins hypothesis was formulated on epidemiological investigations that found a relation between birth and early childhood in poor areas and increased risk and severity of ischemic heart attack in middle age [8]. **B**) Many intervention studies on animal models provide convincing evidence that prenatal conditions affect epigenetic information at CHD candidate loci resulting in phenotypes that lead to CHD development in humans [65,66]. **C**) In Chapter 3 of this thesis we present evidence that prenatal exposure to the Dutch Famine of 1944-1945 affected DNA methylation at CHD candidate loci, but the study was not fit due to size and age of the participants to investigate a further association between DNA methylation and CHD directly. However, epidemiological studies on both Dutch Famine cohorts do indicate the prenatal exposure to it is associated with increased manifestation of the pre-stages of CHD [167,168], although DNA methylation is not indicated in these studies as a potential explanation. **D**) in Chapter 5 we find an association of risk for MI with DNA methylation at 2 of the loci that showed persistent changes in Chapter 3. Although we have no specific information on the early life conditions of the study participants, we do know that they were born in the same period as the populations, that were the basis of the developmental origins hypothesis.

and / or environmentally driven, during the adult period (Chapter 4). Thus, although a developmental explanation for its association with CHD risk is still plausible, it is equally plausible to explain its hypermethylation in cases with epigenetic events in adult life. This can only be resolved with data on both prenatal and adult life conditions of the PROSPER participants.

In **chapter six** we performed a genome-wide investigation on autopsy derived samples from 6 individuals (blood, subcutaneous fat, visceral fat, muscle, liver, pancreas, and spleen) using the recently launched Illumina 450k DNA methylation chip on the capacity of DNA methylation in blood to mark that of internal tissues. Principal component analysis (PCA) revealed that DNA methylation patterns remain stable within at least 24 hours after death, demonstrating that autopsy derived tissue samples can be used to create reference DNA methylomes of inaccessible tissues. PCA further showed that genome-wide variation in DNA methylation is primarily due to tissue differences, dismissing the use of surrogate tissues to mark whole methylome patterns of internal tissues. In contrast, we observed that average DNA methylation levels were similar between blood and each internal tissue at most CpG sites, and across all tissues at roughly a quarter of CpG sites. Both observations were in line with existing literature [100–102].

Chapter 2 showed that at specific loci DNA methylation in blood may correlate with DNA methylation in another tissue, which a recent genome wide study expanded upon [234]. We therefore inspected correlation of CpG methylation at individual sites across the tissues, focusing mostly on CpG rich areas for technical reasons, and found that over 30 % of CpG sites showed a strong correlation ($r \geq 0.75$) between blood and another tissue. At the majority of CpG sites such correlations were between blood and one tissue exclusively, although at a sizeable minority they were found across multiple tissues. At a few CpG sites we observed correlation across all tissues which likely reflects the effect of unknown genetic variation [92,243]. Strong correlation coefficients were not associated with average DNA methylation, its variation, or its difference between the tissues. Our observations were repeated in a recent study on the same dataset extended with additional samples and analyzed using different

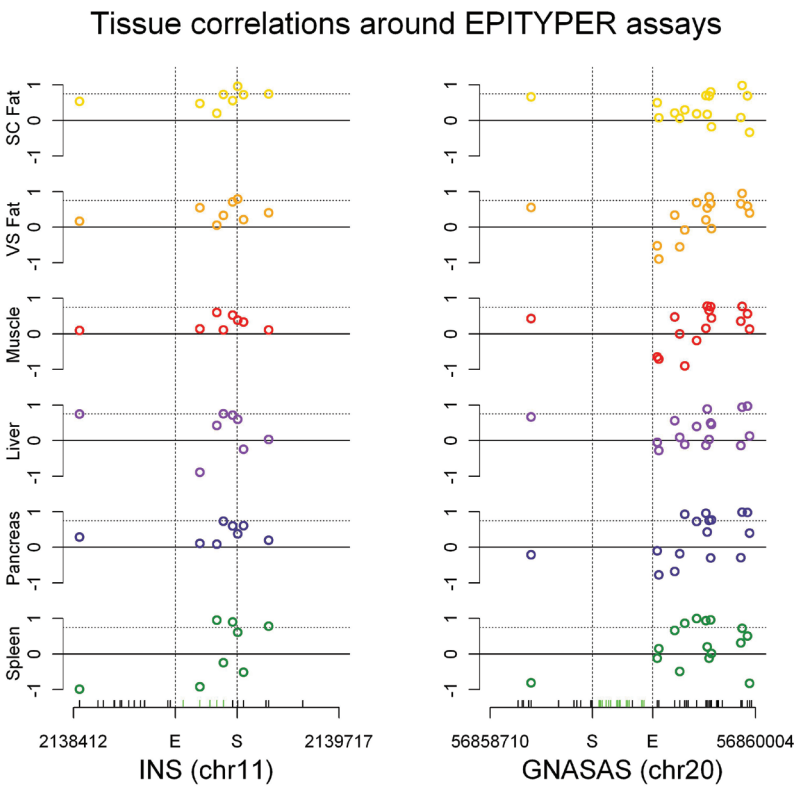


Figure 7.2: Correlation of CpG methylation between blood and spleen (green), pancreas (blue), liver (purple), skeletal muscle (red), visceral fat (orange), and subcutaneous fat (yellow) at CpG sites within and around 500 bp from either start (vertical dotted line at S) or end (vertical dotted line at E) of the SEQUENOM EPITYPER assays for the candidate loci *INS* (left) and *GNASAS* (right) measured in Chapters 2 to 5. The numbers on the x-axis correspond to the chromosomal coordinates (hg 18) of the assay +/- 500 bp, the tick-marks on the x-axis represent all the CpG sites, the green ticks indicate CpG sites of which the methylation was called using the EPITYPER platform. The colored dots represent correlation between blood and the tissues at CpG sites called using the 450K ILLUMINA BeadChip Array in Chapter 6. Horizontal dotted lines are for reference, representing a correlation coefficient of 0.75.

statistical methods [255]. Further, a recent whole methylome study suggested that blood and buccal cells can both be used as surrogate tissues for internal tissues [256]. There are more peripheral tissues that can be obtained with minimal (e.g. buccal cells and epidermis) [107,108] or moderate (e.g. biopsies of subcutaneous fat or muscle) [109,110] invasiveness. As the capacity of a surrogate tissue to mark a target tissue appears locus and tissue specific a reference of DNA methylation correlation across tissues will be of value in choosing the relevant surrogate tissues that mark the target tissue in future epigenetic epidemiological studies.

The current chapter is a pilot when it comes to finding locus specific surrogate tissues, but this data may already help interpret the associations found at *INS* and *GNASAS* in this thesis. We therefore plotted the correlation between blood and the internal tissues for the CpGs on and around ($^{+}/_{-}$ 500 bp) both Sequenom EPITYPER assays (amplicon) measured in chapters 2 to 5 (Figure 7.2). For *INS*, 3 of the CpG sites called in the amplicon were assessed in this chapter, as were 3 CpG site within 60 bp of the Start of the amplicon, and 1 CpG site 470 bp away from the End. The CpG sites around the Start of the *INS* amplicon show some promising correlations between blood and the other tissues. Insulin, under the influence of DNA methylation at *INS*, is produced in pancreatic β-cells [200], which constitute only a tiny fraction of the whole pancreas, yet 1 CpG site shows a correlation with some promise between blood and whole pancreas (Figure 7.2). For *GNASAS* none of the CpG sites called within the amplicon were assessed in this chapter and only 1 CpG site 300 bp from the Start of the amplicon, but within 500 bp from the End of the amplicon we assessed 15 CpG sites in this chapter, of which the CpGs 155 - 455 bp away show promising correlations between blood and other tissues. Higher DNA methylation at *GNASAS* is associated with higher expression of Gαs, a key component of the intracellular cAMP signaling, in adipose tissue (both brown and white), the pituitary and the thyroid gland [201,223]. A few CpG sites show promising correlations between blood and subcutaneous fat or visceral fat. Although there is evidence that small changes in percentage DNA methylation may explain almost half the variance in gene expression [188], and the cAMP and insulin signaling pathways are implicated in metabolic dysregulation [228], it still remains

unclear whether the observed differences associated with MI in chapter 5 indicate an epigenetic process involved in disease development. However, this possibility has not been disproven either, if anything our exploration warrants a bigger investigation with more individuals and samples of tissues involved in cardiovascular and metabolic diseases.

Sequential Study Design

In this thesis we applied a sequence of research questions, each with a specific study design and population cohort, to investigate the causes of epigenetic variation and its consequence for disease risk at the same loci. We found that epigenetic marks can be persistently altered by conditions of the prenatal environment (Chapter 3), that they may change dynamically throughout life (Chapter 4), and that they are associated with risk for CHD (Chapter 5). The summation of random and directed epigenetic variation with a heritable [243], or prenatally determined background fits both the observed randomness in onset of complex diseases, the influence of environmental components on disease development, and the association between conditions during prenatal and early life with risk for these diseases [257]. Unfortunately, there is no single study design that can readily isolate all components of epigenetic variation, nor will one population cohort likely accommodate all research questions involved in investigating the accumulated causes and consequences of epigenetic variation across a lifetime. However, careful formulation of specific research question and subsequent selection of appropriate study designs, may still promise meaningful epigenetic research on ageing and its related diseases [112], as demonstrated by sequentially carrying results across chapters 3, 4, and 5 in this thesis.

The sequential study design that we applied with chapters 3 to 5 combines a retrospective, a (mainly) cross-sectional, and a prospective study, developing epigenetic markers of a past environmental exposure, establishing their susceptibility to epigenetic drift, and using these markers to investigate disease risk in a clinical trial, respectively. This approach exploits the potential of epigenetic research, covering time periods and events from preconception till death that are beyond single population studies and providing insight into

the environmental and stochastic components of the common age related diseases. However, prospective and retrospective investigation necessitates using epigenetic marks with a stable variation for at least the duration of the study. Investigating longitudinal samples would be best for this [49], but, when not available, another form of evidence on temporal stability can be obtained. Similarly, isolating specific leukocyte cell types will be impractical for large studies, many of which even lack information on the proportions of the major leukocyte sub-populations (e.g. chapters 3 and 5), although in genome wide studies another method can be applied [104]. As, in humans, direct analysis of internal tissues is practically impossible in anything but a study on post mortem biosamples, establishing the correlation of methylation in peripheral tissues with these disease relevant tissues will be crucial for deeper interpretation of uncovered associations. Thus, effective epigenetic research requires information on the stability of epigenetic variation over time and its correspondence across tissues and cell types.

In this thesis we used a preliminary sub-study on longitudinal biosamples from healthy individuals (potential controls) to assess the amount of epigenetic variation at the candidate loci, the influence of leukocyte population heterogeneity and its temporal stability during a period covering the follow-up period of most clinical studies (chapter 2). As an auxiliary we used a (small) genome wide study on autopsy samples (Chapter 6) to interpret the associations of *GNASAS* and *INS* with disease risk in relation to the mechanism of disease development. With the eventual completion of epigenetic data bases, such as those of the International Human Epigenome Consortium [96] (http://www.ihec-epigenomes.org/) and the National Institutes of Health (NIH) Epigenomics Roadmap [97] (http://www.roadmapepigenomics.org/), such steps in the sequential design can be performed by a database search. Thus, from this thesis it follows that the design and interpretation of epigenetic associations studies into ageing related diseases requires data bases that record data from: **1)** studies on genome wide DNA methylation marks in longitudinal samples of peripheral tissues from healthy individuals to assess temporal stability (Chapter 2; [88,89]. **2)** studies on postpartum samples of extra-embryonic tissues from neonates [62] to acquire references for developmental DNA

methylation dynamics in humans from gametogenesis till birth; **3)** studies on post-mortem samples of peripheral and internal tissues to assess the influence of cell type population heterogeneity on DNA methylation marks for each tissue (Chapters 2 and 4), and the correlation of methylation marks across tissues (Chapter 6), and preferably to also provide a record of whole DNA methylomes from each cell type [104]; **4)** studies on mono-cultures of multiple cell types to assess the overlap between DNA methylation and the other layers of epigenetic information (i.e. DNA methylation and other base modifications, histone modifications, association with nuclear lamina, or CTCF hubs); **5**) studies on multiple cell cultures coupling epigenomic with transcriptomic, and perhaps also proteomic, analyses to assess the downstream effects of epigenetic signatures on cellular gene expression; **6**) classical twin studies comparing monozygotic, dizygotic twin pairs and unrelated individuals to assess the influence of heritable genetic sequence variation and environmental (prenatal and late) factors on the epigenetic status at a locus; and **7**) human intervention studies to investigate the epigenetic consequences of medication, therapy and lifestyle changes.

Statistical analyses

Studying the epigenetic basis of common disease not only requires specific study designs, but also particular statistical tools. We implemented linear mixed models to address issues for analyzing DNA methylation data. It uses all available methylation data, accounts for the correlated CpG methylation of a locus, actually using this information to handle data missing at random (Chapters 2, 3, 4, and 5) and achieve accurate imputation for further analyses such as logistic regression (Chapter 5). Besides powerful, the linear mixed model is also versatile and we used several adaptations of it in our various analyses throughout this thesis. The model simultaneously tests whether the observed effect is independent of included covariates (e.g. age, sex, blood parameters, exposures) and adjusts the effect for it. We used this aspect to make all relevant adjustments on the raw data simultaneously (Chapter 2, 3, 4, and 5). The model also calculates the amount of variation that remains unaccounted for by its covariate structure (i.e. residual variance). We used this residual variance in a nested model approach [156]

to assess the methylation variation explained by each of the major leukocyte sub types (Chapters 2 and 4). Random effect covariates allow estimation of changes in variation, which we used to investigate ageing related changes in methylation variation (Chapter 4). Further, by inclusion of multiple subject variables in the model, we assessed the separate effects of shared and unique environmental factors on age related variation in DNA methylation (Chapter 4). The crucial characteristic of a statistical model for analyzing CpG methylation at candidate loci is that it can accommodate the correlation in methylation between adjacent CpG sites (Chapter 2). The linear mixed model can, and is our model of choice for this type of data due to its statistical robustness and its analytical versatility. However, other models that can handle the correlation of adjacent CpG sites exist and can be used.

Future prospects

In this thesis we explored the limits and possibilities of the emerging research field of epigenetic epidemiology, investigating the kind of research questions on the developmental and environmental components of CHD that existing biobanks can be used for. We first described characteristics of variation in DNA methylation at 16 known CHD candidate loci with demonstrated or suspected epigenetic regulation. Then, using a sequence of studies with linked design, we found evidence that epigenetic signatures at some of these loci can persistently change due to environmental (prenatal) conditions, are susceptible to stochastic perturbations that gradually accumulate during life and that epigenetic signatures at two of these loci were associated with risk for CHD. Expanding on these results, epigenetic research promises the potential to reveal the molecular mechanisms that underlie the influence of the developmental period in CHD, the environmental components of CHD and the randomness in CHD development [112,257].

The epigenetic epidemiology of CHD and other complex diseases is still in its infancy, but a recent study reported an association of adult smoking with DNA methylation signatures persisting in blood years after the exposure [71] and in a second study of linked design the same marks were found

associated with risk for CHD [258]. Such results parallel results from the sequential study design in this thesis and present another model of a disease related exposure that is associated with epigenetic marks that are also associated with the outcome. Whether the associated epigenetic signatures mark the exposure or mediate between exposure and outcome remains to be revealed. Interestingly, another study reported DNA methylation associated with prenatal exposure to smoking [67]. In addition there is increasing speculation of transgenerational epigenetic inheritance, with some associations reported between environmental exposures and phenotypic changes across multiple generations [259,260]. However, no convincing empirical evidence has yet been reported on its importance for humans in context of CHD development [112].

Current epigenetic knowledge may help understand the functional relevance of major GWAS hits in non-coding genomic regions [252], but it only scratches at the surface of the epigenetic contribution to CHD development in later life. Assessing the phenotypic relevance of associations between epigenetic marks and CHD in this thesis and other studies [55,254] requires a much deeper understanding of the epigenome, the origin of epigenetic variation, and its causal relation with disease development. A technological hurdle to be cleared for epigenetic epidemiology is the creation of a platform that integrates the possibilities of whole genome bisulphite sequencing [83], with the simultaneous genome wide calling of other modified bases, such as hmC, and various histone modifications [261,262]. Such integrated whole epigenome scans will allow a better assessment of the function of disease associated epigenetic marks in relation to all other epigenetic marks at the locus and throughout the epigenome. A practical hurdle is the necessity to use peripheral tissues as surrogates for inaccessible tissues that are relevant for disease development. Current research has started to map the possibilities of epigenetic marks in accessible tissues for use as surrogate markers [255,256], but the potentially interesting epigenetic signatures of some tissues will likely remain out of reach. Methodological hurdles to be cleared are the establishment of an atlas of soma-wide epigenetic variation as reference [96,97] development of study designs that enable investigating the mechanisms behind the establishment and maintenance of epigenetic marks and the environmental factors that affect

these processes from preconception throughout life [15,251,263]; and the creation of an integrated functional –omics paradigm that combines genomic, epigenomic, and transcriptomic data [47,112]. Eventually such knowledge will aid understanding the downstream effects of disease associated epigenetic marks on gene expression of the locus in relation to cellular expression pathways, and the functionality of the tissue(s) involved in the process of disease development [94,263]

References

1. Vaartjes I, Koopman C, van Dis I, Visseren FLJ, and Bots ML (2013) Hart- en vaatziekten in Nederland. In: Vaartjes I, Koopman C, van Dis I, Visseren FLJ, Bots ML, editors. Hart- en vaatziekten in Nederland, 2013, cijfers over leefstijl, risicofactoren, ziekte en sterfte. Den Haag: Hartstichting. pp. 7-26.

2. Dermitzakis ET and Clark AG (2009) Genetics. Life after GWA studies. Science 326: 239-240.

3. Mechanic LE, Chen HS, Amos CI, Chatterjee N, Cox NJ, Divi RL, et al. (2011) Next generation analytic tools for large scale genetic epidemiology studies of complex diseases. Genet Epidemiol .

4. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, and de Faire U (2002) Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. J Intern Med 252: 247-254.

5. Brunham LR and Hayden MR (2013) Hunting human disease genes: lessons from the past, challenges for the future. Hum Genet 132: 603-617.

6. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, and Kannel WB (1998) Prediction of coronary heart disease using risk factor categories. Circulation 97: 1837-1847.

7. Graham I, Atar D, Borch-Johnsen K, Boysen G, Burell G, Cifkova R, et al. (2007) European guidelines on cardiovascular disease prevention in clinical practice: executive summary: Fourth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (Constituted by representatives of nine societies and by invited experts). Eur Heart J 28: 2375-2414.

8. Barker DJ and Osmond C (1986) Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. Lancet 1: 1077-1081.

9. Leon DA, Lithell HO, Vagero D, Koupilova I, Mohsen R, Berglund L, Lithell UB, and McKeigue PM (1998) Reduced fetal growth rate and increased risk of death from ischaemic heart disease: cohort study of 15 000 Swedish men and women born 1915-29. BMJ 317: 241-245.

10. Lumey LH, Stein AD, and Susser E (2011) Prenatal famine and adult health. Annu Rev Public Health 32: 237-262.

11. Lumey LH, Stein AD, Kahn HS, van der Pal-de Bruin KM, Blauw G, Zybert PA, and Susser ES (2007) Cohort Profile: the Dutch Hunger Winter Families Study. Int J Epidemiol 36: 1196-1204.

12. Weaver IC (2007) Epigenetic programming by maternal behavior and pharmacological intervention. Nature versus nurture: let's call the whole thing off. Epigenetics 2: 22-28.

13. Bogdarina I, Welham S, King PJ, Burns SP, and Clark AJ (2007) Epigenetic modification of the renin-angiotensin system in the fetal programming of hypertension. Circ Res 100: 520-526.

14. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, and Lumey LH (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci U S A 105: 17046-17049.

15. Feil R and Fraga MF (2011) Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 13: 97-109.

16. Winnefeld M and Lyko F (2012) The aging epigenome: DNA methylation from the cradle to the grave. Genome Biol 13: 165.

17. Feinberg AP (2008) Epigenetics at the epicenter of modern medicine. JAMA 299: 1345-1350.

18. Feinberg AP (2013) The epigenetic basis of common human disease. Trans Am Clin Climatol Assoc 124: 84-93.

19. Tremethick DJ (2007) Higher-order structures of chromatin: the elusive 30 nm fiber. Cell 128: 651-654.

20. Misteli T (2007) Beyond the sequence: cellular organization of genome function. Cell 128: 787-800.

21. Zhou VW, Goren A, and Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet 12: 7-18.

22. Kouzarides T (2007) Chromatin modifications and their function. Cell 128: 693-705.

23. Bernstein BE, Meissner A, and Lander ES (2007) The mammalian epigenome. Cell 128: 669-681.

24. Suzuki MM and Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465-476.

25. Cedar H and Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10: 295-304.

26. Li B, Carey M, and Workman JL (2007) The role of chromatin during transcription. Cell 128: 707-719.

27. Bestor TH (2003) Cytosine methylation mediates sexual conflict. Trends Genet 19: 185-190.

28. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13: 484-492.

29. Jaenisch R and Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33 Suppl: 245-254.

30. Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447: 425-432.

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

32. Ferguson-Smith AC (2011) Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet 12: 565-575.

33. Plagge A and Kelsey G (2006) Imprinting the Gnas locus. Cytogenet Genome Res 113: 178-187.

34. Constancia M, Kelsey G, and Reik W (2004) Resourceful imprinting. Nature 432: 53-57.

35. Weaver JR, Susiarjo M, and Bartolomei MS (2009) Imprinting and epigenetic changes in the early embryo. Mamm Genome 20: 532-543.

36. Shemer R, Hershko AY, Perk J, Mostoslavsky R, Tsuberi B, Cedar H, Buiting K, and Razin A (2000) The imprinting box of the Prader-Willi/Angelman syndrome domain. Nat Genet 26: 440-443.

37. Smith AC, Choufani S, Ferreira JC, and Weksberg R (2007) Growth regulation, imprinted genes, and chromosome 11p15.5. Pediatr Res 61: 43R-47R.

38. Hashimoto H, Vertino PM, and Cheng X (2010) Molecular coupling of DNA methylation and histone methylation. Epigenomics 2: 657-669.

39. Miranda TB and Jones PA (2007) DNA methylation: the nuts and bolts of repression. J Cell Physiol 213: 384-390.

40. Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, and Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97: 5237-5242.

41. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315-322.

42. Kriaucionis S and Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324: 929-930.

43. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, and Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324: 930-935.

44. Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, and Sowers LC (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res 32: 4100-4108.

45. Branco MR, Ficz G, and Reik W (2012) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet 13: 7-13.

46. Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, Bruckl T, Biel M, and Carell T (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE 5: e15367.

47. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74.

48. Kuehnen P, Mischke M, Wiegand S, Sers C, Horsthemke B, Lau S, Keil T, Lee YA, Grueters A, and Krude H (2012) An Alu element-associated hypermethylation variant of the POMC gene is associated with childhood obesity. PLoS Genet 8: e1002543.

49. Toperoff G, Aran D, Kark JD, Rosenberg M, Dubnikov T, Nissan B, Wainstein J, Friedlander Y, Levy-Lahad E, Glaser B, and Hellman A (2012) Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. Hum Mol Genet 21: 371-383.

50. Issa JP (2011) Epigenetic variation and cellular Darwinism. Nat Genet 43: 724-726.

51. Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R, et al. (2012) Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat Genet 44: 1207-1214.

52. Jones PA and Baylin SB (2002) The fundamental role of epigenetic events in cancer. Nat Rev Genet 3: 415-428.

53. Jones PA and Baylin SB (2007) The epigenomics of cancer. Cell 128: 683-692.

54. Baylin SB and Jones PA (2011) A decade of exploring the cancer epigenome - biological and translational implications. Nat Rev Cancer 11: 726-734.

55. Movassagh M, Choy MK, Goddard M, Bennett MR, Down TA, and Foo RS (2010) Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. PLoS ONE 5: e8564.

56. Kaneda R, Takada S, Yamashita Y, Choi YL, Nonaka-Sarukawa M, Soda M, Misawa Y, Isomura T, Shimada K, and Mano H (2009) Genome-wide histone methylation profile for heart failure. Genes Cells 14: 69-77.

57. Ollikainen M and Craig JM (2011) Epigenetic discordance at imprinting control regions in twins. Epigenomics 3: 295-306.

58. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet 8: e1002629.

59. Martin GM (2009) Epigenetic gambling and epigenetic drift as an antagonistic pleiotropic mechanism of aging. Aging Cell 8: 761-764.

60. Heijmans BT, Kremer D, Tobi EW, Boomsma DI, and Slagboom PE (2007) Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. Hum Mol Genet 16: 547-554.

61. Gertz J, Varley KE, Reddy TE, Bowling KM, Pauli F, Parker SL, Kucera KS, Willard HF, and Myers RM (2011) Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. PLoS Genet 7: e1002228.

62. Ollikainen M, Smith KR, Joo EJ, Ng HK, Andronikos R, Novakovic B, bdul Aziz NK, Carlin JB, Morley R, Saffery R, and Craig JM (2010) DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Hum Mol Genet 19: 4176-4188.

63. Dolinoy DC, Weidman JR, Waterland RA, and Jirtle RL (2006) Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. Environ Health Perspect 114: 567-572.

64. Goldberg AD, Allis CD, and Bernstein E (2007) Epigenetics: a landscape takes shape. Cell 128: 635-638.

65. Gluckman PD, Hanson MA, Buklijas T, Low FM, and Beedle AS (2009) Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. Nat Rev Endocrinol 5: 401-408.

66. Attig L, Gabory A, and Junien C (2010) Early nutrition and epigenetic programming: chasing shadows. Curr Opin Clin Nutr Metab Care 13: 284-293.

67. Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, et al. (2012) 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect 120: 1425-1431.

68. Jirtle RL and Skinner MK (2007) Environmental epigenomics and disease susceptibility. Nat Rev Genet 8: 253-262.

69. Heijmans BT, Tobi EW, Lumey LH, and Slagboom PE (2009) The epigenome: Archive of the prenatal environment. Epigenetics 4: 526-531.

70. Petronis A (2006) Epigenetics and twins: three variations on the theme. Trends Genet 22: 347-350.

71. Breitling LP, Yang R, Korn B, Burwinkel B, and Brenner H (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet 88: 450-457.

72. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, and Philibert RA (2012) Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. Am J Med Genet B Neuropsychiatr Genet 159B: 141-151.

73. Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, Byun HM, Jiang J, Marinelli B, Pesatori AC, Bertazzi PA, and Yang AS (2007) Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res 67: 876-880.

74. Ren X, McHale CM, Skibola CF, Smith AH, Smith MT, and Zhang L (2011) An emerging role for epigenetic dysregulation in arsenic toxicity and carcinogenesis. Environ Health Perspect 119: 11-19.

75. Suarez-Alvarez B, Baragano RA, Ortega F, and Lopez-Larrea C (2013) Epigenetic modulation of the immune function: A potential target for tolerance. Epigenetics 8.

76. Kubo M and Motomura Y (2012) Transcriptional regulation of the anti-inflammatory cytokine IL-10 in acquired immune cells. Front Immunol 3: 275.

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

78. Harrison A and Parle-McDermott A (2011) DNA methylation: a timeline of methods and applications. Front Genet 2: 74.

79. Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, and Meissner A (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol 28: 1106-1114.

80. Novak P, Stampfer MR, Munoz-Rodriguez JL, Garbe JC, Ehrich M, Futscher BW, and Jensen TJ (2012) Cell-type specific DNA methylation patterns define human breast cellular identity. PLoS ONE 7: e52299.

81. Jin SG, Kadam S, and Pfeifer GP (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res 38: e125.

82. Gupta R, Nagarajan A, and Wajapeyee N (2010) Advances in genome-wide DNA methylation analysis. Biotechniques 49: iii-ixi.

83. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. (2012) Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci U S A 109: 10522-10527.

84. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. (2006) High-throughput DNA methylation profiling using universal bead arrays. Genome Res 16: 383-393.

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

86. Colella S, Shen L, Baggerly KA, Issa JP, and Krahe R (2003) Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. Biotechniques 35: 146-150.

87. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, and van den Boom D (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A 102: 15785-15790.

88. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766-770.

89. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473: 43-49.

90. Bock C, Walter J, Paulsen M, and Lengauer T (2008) Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping. Nucleic Acids Res 36: e55.

91. McDaniell R, Lee BK, Song L, Liu Z, Boyle AP, Erdos MR, et al. (2010) Heritable individual-specific and allele-specific chromatin signatures in humans. Science 328: 235-239.

92. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, and Pritchard JK (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol 12: R10.

93. Weaver IC, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, Dymov S, Szyf M, and Meaney MJ (2004) Epigenetic programming by maternal behavior. Nat Neurosci 7: 847-854.

94. Rakyan VK, Down TA, Balding DJ, and Beck S (2011) Epigenome-wide association studies for common human diseases. Nat Rev Genet 12: 529-541.

95. Xie H, Wang M, de Andrade A, Bonaldo MD, Galat V, Arndt K, Rajaram V, Goldman S, Tomita T, and Soares MB (2011) Genome-wide quantitative assessment of variation in DNA methylation patterns. Nucleic Acids Res .

96. Jones PA, Archer TK, Baylin SB, Beck S, Berger S, Bernstein BE, et al. (2008) Moving AHEAD with an international human epigenome project. Nature 454: 711-715.

97. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. (2010) The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 28: 1045-1048.

98. Foley DL, Craig JM, Morley R, Olsson CJ, Dwyer T, Smith K, and Saffery R (2009) Prospects for epigenetic epidemiology. Am J Epidemiol 169: 389-400.

99. Relton CL and Davey SG (2010) Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. PLoS Med 7: e1000356.

100. Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, Laird PW, and Yang AS (2009) Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. Hum Mol Genet 18: 4808-4817.

101. Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, and Held WA (2005) Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A 102: 3336-3341.

102. Illingworth R, Kerr A, Desousa D, Jorgensen H, Ellis P, Stalker J, et al. (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol 6: e22.

103. Martin GM (2005) Epigenetic drift in aging identical twins. Proc Natl Acad Sci U S A 102: 10413-10414.

104. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, and Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13: 86.

105. McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonte B, Szyf M, Turecki G, and Meaney MJ (2009) Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nat Neurosci 12: 342-348.

106. Hoyo C, Murphy SK, and Jirtle RL (2009) Imprint regulatory elements as epigenetic biosensors of exposure in epidemiological studies. J Epidemiol Community Health 63: 683-684.

107. Freeman B, Smith N, Curtis C, Huckett L, Mill J, and Craig IW (2003) DNA from buccal swabs recruited by mail: evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. Behav Genet 33: 67-72.

108. Gronniger E, Weber B, Heil O, Peters N, Stab F, Wenck H, Korn B, Winnefeld M, and Lyko F (2010) Aging and chronic sun exposure cause distinct epigenetic changes in human skin. PLoS Genet 6: e1000971.

109. Noer A, Sorensen AL, Boquest AC, and Collas P (2006) Stable CpG hypomethylation of adipogenic promoters in freshly isolated, cultured, and differentiated mesenchymal stem cells from adipose tissue. Mol Biol Cell 17: 3543-3556.

110. Ronn T, Poulsen P, Hansson O, Holmkvist J, Almgren P, Nilsson P, Tuomi T, Isomaa B, Groop L, Vaag A, and Ling C (2008) Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle. Diabetologia 51: 1159-1168.

111. Murphy SK, Huang Z, and Hoyo C (2012) Differentially methylated regions of imprinted genes in prenatal, perinatal and postnatal human tissues. PLoS ONE 7: e40924.

112. Mill J and Heijmans BT (2013) From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet 14: 585-594.

113. Boomsma DI, de Geus EJ, Vink JM, Stubbe JH, Distel MA, Hottenga JJ, Posthuma D, van Beijsterveldt TC, Hudziak JJ, Bartels M, and Willemsen G (2006) Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet 9: 849-857.

114. Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, et al. (2008) Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. Eur J Hum Genet 16: 335-342.

115. Wang L, Wang F, Guan J, Le J, Wu L, Zou J, Zhao H, Pei L, Zheng X, and Zhang T (2010) Relation between hypomethylation of long interspersed nucleotide elements and risk of neural tube defects. Am J Clin Nutr 91: 1359-1367.

116. Willemsen G, de Geus EJ, Bartels M, van Beijsterveldt CE, Brooks AI, Estourgie-van Burk GF, et al. (2010) The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. Twin Res Hum Genet 13: 231-245.

117. McGue M and Christensen K (2007) Social activity and healthy aging: a study of aging Danish twins. Twin Res Hum Genet 10: 255-265.

118. Shepherd J, Blauw GJ, Murphy MB, Cobbe SM, Bollen EL, Buckley BM, et al. (1999) The design of a prospective study of Pravastatin in the Elderly at Risk (PROSPER). PROSPER Study Group. PROspective Study of Pravastatin in the Elderly at Risk. Am J Cardiol 84: 1192-1197.

119. Waterland RA and Michels KB (2007) Epigenetic Epidemiology of the Developmental Origins Hypothesis. Annu Rev Nutr .

120. Steegers-Theunissen RP, Obermann-Borst SA, Kremer D, Lindemans J, Siebel C, Steegers EA, Slagboom PE, and Heijmans BT (2009) Periconceptional maternal folic acid use of 400 µg per day is related to increased methylation of the IGF2 gene in the very young child. PLoS ONE 4: e7845.

121. Kouzarides T (2007) SnapShot: Histone-modifying enzymes. Cell 128: 802-U27.

122. Petronis A (2001) Human morbid genetics revisited: relevance of epigenetics. Trends Genet 17: 142-146.

123. Pons D, de Vries FR, van den Elsen PJ, Heijmans BT, Quax PH, and Jukema JW (2009) Epigenetic histone acetylation modifiers in vascular remodelling: new targets for therapy in cardiovascular disease. Eur Heart J 30: 266-277.

124. Turunen MP, Aavik E, and Yla-Herttuala S (2009) Epigenetics and atherosclerosis. Biochim Biophys Acta .

125. Shen L and Waterland RA (2007) Methods of DNA methylation analysis. Curr Opin Clin Nutr Metab Care 10: 576-581.

126. Smith ZD, Gu H, Bock C, Gnirke A, and Meissner A (2009) High-throughput bisulfite sequencing in mammalian genomes. Methods 48: 226-232.

127. Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox AV, et al. (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. PLoS Biol 2: e405.

128. Waterland RA and Jirtle RL (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol 23: 5293-5300.

129. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 41: 178-186.

130. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102: 10604-10609.

131. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, et al. (2008) Intra-individual change over time in DNA methylation with familial clustering. JAMA 299: 2877-2883.

132. Ito Y, Koessler T, Ibrahim AE, Rai S, Vowler SL, bu-Amero S, et al. (2008) Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer. Hum Mol Genet 17: 2633-2643.

133. Cui H, Cruz-Correa M, Giardiello FM, Hutcheon DF, Kafonek DR, Brandenburg S, Wu Y, He X, Powe NR, and Feinberg AP (2003) Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. Science 299: 1753-1755.

134. Ally MS, Al-Ghnaniem R, and Pufulete M (2009) The relationship between gene-specific DNA methylation in leukocytes and normal colorectal mucosa in subjects with and without colorectal tumors. Cancer Epidemiol Biomarkers Prev 18: 922-928.

135. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, Slagboom PE, and Heijmans BT (2009) DNA Methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum Mol Genet 18: 4046-4053.

136. Dong J, Ivascu C, Chang HD, Wu P, Angeli R, Maggi L, et al. (2007) IL-10 Is Excluded from the Functional Cytokine Memory of Human CD4+ Memory T Lymphocytes. J Immunol 179: 2389-2396.

137. Hayward BE, Kamiya M, Strain L, Moran V, Campbell R, Hayashizaki Y, and Bonthron DT (1998) The human GNAS1 gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. Proc Natl Acad Sci U S A 95: 10038-10043.

138. Sullivan KE, Reddy AB, Dietzmann K, Suriano AR, Kocieda VP, Stewart M, and Bhatia M (2007) Epigenetic regulation of tumor necrosis factor alpha. Mol Cell Biol 27: 5147-5160.

139. Sandovici I, Leppert M, Hawk PR, Suarez A, Linares Y, and Sapienza C (2003) Familial aggregation of abnormal methylation of parental alleles at the IGF2/H19 and IGF2R differentially methylated regions. Hum Mol Genet 12: 1569-1578.

140. Arnaud P, Monk D, Hitchins M, Gordon E, Dean W, Beechey CV, Peters J, Craigen W, Preece M, Stanier P, Moore GE, and Kelsey G (2003) Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. Hum Mol Genet 12: 1005-1019.

141. Melzner I, Scott V, Dorsch K, Fischer P, Wabitsch M, Bruderlein S, Hasel C, and Moller P (2002) Leptin gene expression in human preadipocytes is switched on by maturation-induced demethylation of distinct CpGs in its proximal promoter. J Biol Chem 277: 45420-45427.

142. McGill BE, Bundle SF, Yaylaoglu MB, Carson JP, Thaller C, and Zoghbi HY (2006) Enhanced anxiety and stress-induced corticosterone release are associated with increased Crh expression in a mouse model of Rett syndrome. Proc Natl Acad Sci U S A 103: 18267-18272.

143. Probst MC, Thumann H, Aslanidis C, Langmann T, Buechler C, Patsch W, Baralle FE, linga-Thie GM, Geisel J, Keller C, Menys VC, and Schmitz G (2004) Screening for functional sequence variations and mutations in ABCA1. Atherosclerosis 175: 269-279.

144. Adkins RM, Fain JN, Krushkal J, Klauser CK, Magann EF, and Morrison JC (2007) Association between paternally inherited haplotypes upstream of the insulin gene and umbilical cord IGF-II levels. Pediatr Res 62: 451-455.

145. Mitsuya K, Meguro M, Lee MP, Katoh M, Schulz TC, Kugoh H, Yoshida MA, Niikawa N, Feinberg AP, and Oshimura M (1999) LIT1, an imprinted antisense RNA in the human KvLQT1 locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. Hum Mol Genet 8: 1209-1217.

146. Rosa AL, Wu YQ, Kwabi-Addo B, Coveler KJ, Sutton VR, and Shaffer LG (2005) Allele-specific methylation of a functional CTCF binding site upstream of MEG3 in the human imprinted domain of 14q32. Chromosome Res 13: 809-818.

147. Medstrand P, Landry JR, and Mager DL (2001) Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. J Biol Chem 276: 1896-1903.

148. Hayward BE and Bonthron DT (2000) An imprinted antisense transcript at the human GNAS1 locus. Hum Mol Genet 9: 835-841.

149. Li LC and Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. Bioinformatics 18: 1427-1431.

150. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, and Haussler D (2002) The human genome browser at UCSC. Genome Res 12: 996-1006.

151. Coolen MW, Statham AL, Gardiner-Garden M, and Clark SJ (2007) Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res 35: e119.

152. Ehrich M, Turner J, Gibbs P, Lipton L, Giovanneti M, Cantor C, and van den Boom D (2008) Cytosine methylation profiling of cancer cell lines. Proc Natl Acad Sci U S A 105: 4844-4849.

153. Thompson RF, Suzuki M, Lau KW, and Greally JM (2009) A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry. Bioinformatics 25: 2164-2170.

154. Stanssens P, Zabeau M, Meersseman G, Remes G, Gansemans Y, Storm N, Hartmer R, Honisch C, Rodi CP, Bocker S, and van den Boom D (2004) High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. Genome Res 14: 126-133.

155. Altman DG (1991) Practical statistics for medical research. London; New York: Chapman and Hall.

156. West B, Welch K, and Galecki A (12-1-2006) Linear Mixed Models: A Practical Guide Using Statistical Software. Chapmann Hall / CRC Press.

157. El-Maarri O, Becker T, Junen J, Manzoor SS, az-Lacava A, Schwaab R, Wienker T, and Oldenburg J (2007) Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. Hum Genet 122: 505-514.

158. Herman JG, Graff JR, Myohanen S, Nelkin BD, and Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci U S A 93: 9821-9826.

159. The International HapMap Consortium (2003) The International HapMap Project. Nature 426: 789-796.

160. Bourc'his D, Xu GL, Lin CS, Bollman B, and Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294: 2536-2539.

161. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, and Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448: 714-717.

162. Ramchandani S, Bhattacharya SK, Cervoni N, and Szyf M (1999) DNA methylation is a reversible biological signal. Proc Natl Acad Sci U S A 96: 6107-6112.

163. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P, and Baccarelli A (2009) Decline in genomic DNA methylation through aging in a cohort of elderly subjects. Mech Ageing Dev 130: 234-239.

164. Hales CN and Barker DJ (2001) The thrifty phenotype hypothesis. Br Med Bull 60: 5-20.

165. Barker DJ (2007) The origins of the developmental origins theory. J Intern Med 261: 412-417.

166. Whincup PH, Kaye SJ, Owen CG, Huxley R, Cook DG, Anazawa S, et al. (2008) Birth weight and risk of type 2 diabetes: a systematic review. JAMA 300: 2886-2897.

167. Lumey LH, Stein AD, Kahn HS, and Romijn JA (2009) Lipid profiles in middle-aged men and women after famine exposure during gestation: the Dutch Hunger Winter Families Study. Am J Clin Nutr 89: 1737-1743.

168. Roseboom TJ, de Rooij S, and Painter R (2006) The Dutch famine and its long-term consequences for adult health. Early Hum Dev 82: 485-491.

169. Stein AD, Kahn HS, Rundle A, Zybert PA, van der Pal-de Bruin, and Lumey LH (2007) Anthropometric measures in middle age after exposure to famine during gestation: evidence from the Dutch famine. Am J Clin Nutr 85: 869-876.

170. Susser E, Neugebauer R, Hoek HW, Brown AS, Lin S, Labovitz D, and Gorman JM (1996) Schizophrenia after prenatal famine. Further evidence. Arch Gen Psychiatry 53: 25-31.

171. Kyle UG and Pichard C (2006) The Dutch Famine of 1944-1945: a pathophysiological model of long-term consequences of wasting disease. Curr Opin Clin Nutr Metab Care 9: 388-394.

172. Brown AS and Susser ES (1997) Sex differences in prevalence of congenital neural defects after periconceptional famine exposure. Epidemiology 8: 55-58.

173. Ravelli AC, van der Meulen JH, Osmond C, Barker DJ, and Bleker OP (1999) Obesity at the age of 50 y in men and women exposed to famine prenatally. Am J Clin Nutr 70: 811-816.

174. Bygren LO, Edvinsson S, and Brostrom G (2000) Change in food availability during pregnancy: Is it related to adult sudden death from cerebro- and cardiovascular disease in offspring? Am J Hum Biol 12: 447-453.

175. Painter RC, de Rooij S, Bossuyt PM, Simmers TA, Osmond C, Barker DJ, Bleker OP, and Roseboom TJ (2006) Early onset of coronary artery disease after prenatal exposure to the Dutch famine. Am J Clin Nutr 84: 322-327.

176. Gilbert JS and Nijland MJ (2008) Sex differences in the developmental origins of hypertension and cardiorenal disease. Am J Physiol Regul Integr Comp Physiol 295: R1941-R1952.

177. Nathanielsz PW (2006) Animal models that elucidate basic principles of the developmental origins of adult diseases. ILAR J 47: 73-82.

178. Gluckman PD, Hanson MA, Cooper C, and Thornburg KL (2008) Effect of in utero and early-life conditions on adult health and disease. N Engl J Med 359: 61-73.

179. Waterland RA, Dolinoy DC, Lin JR, Smith CA, Shi X, and Tahiliani KG (2006) Maternal methyl supplements increase offspring DNA methylation at Axin Fused. Genesis 44: 401-406.

180. Monk D, Sanches R, Arnaud P, Apostolidou S, Hills FA, bu-Amero S, Murrell A, Friess H, Reik W, Stanier P, Constancia M, and Moore GE (2006) Imprinting of IGF2 P0 transcript and novel alternatively spliced INS-IGF2 isoforms show differences between mouse and human. Hum Mol Genet 15: 1259-1269.

181. Morgan HD, Santos F, Green K, Dean W, and Reik W (2005) Epigenetic reprogramming in mammals. Hum Mol Genet 14 Spec No 1: R47-R58.

182. Reik W, Dean W, and Walter J (2001) Epigenetic reprogramming in mammalian development. Science 293: 1089-1093.

183. Morgan HD, Jin XL, Li A, Whitelaw E, and O'Neill C (2008) The culture of zygotes to the blastocyst stage changes the postnatal expression of an epigentically labile allele, agouti viable yellow, in mice. Biol Reprod 79: 618-623.

184. Sinclair KD, Allegrucci C, Singh R, Gardner DS, Sebastian S, Bispham J, et al. (2007) DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status. Proc Natl Acad Sci U S A 104: 19351-19356.

185. Yamagata Y, Asada H, Tamura I, Lee L, Maekawa R, Taniguchi K, Taketani T, Matsuoka A, Tamura H, and Sugino N (2009) DNA methyltransferase expression in the human endometrium: down-regulation by progesterone and estrogen. Hum Reprod 24: 1126-1132.

186. Bjornsson HT, Fallin MD, and Feinberg AP (2004) An integrated epigenetic and genetic approach to common human disease. Trends Genet 20: 350-358.

187. Duijts L, Bakker-Jonges LE, Labout JA, Jaddoe VW, Hofman A, Steegers EA, van Dongen JJ, Hooijkaas H, and Moll HA (2009) Fetal growth influences lymphocyte subset counts at birth: the Generation R Study. Neonatology 95: 149-156.

188. Lillycrop KA, Phillips ES, Torrens C, Hanson MA, Jackson AA, and Burdge GC (2008) Feeding pregnant rats a protein-restricted diet persistently alters the methylation of specific cytosines in the hepatic PPAR alpha promoter of the offspring. Br J Nutr 100: 278-282.

189. Stöger R (2008) The thrifty epigenotype: an acquired and heritable predisposition for obesity and diabetes? Bioessays 30: 156-166.

190. Prentice AM, Rayco-Solon P, and Moore SE (2005) Insights from the developing world: thrifty genotypes and thrifty phenotypes. Proc Nutr Soc 64: 153-161.

191. Stoger R (2006) In vivo methylation patterns of the leptin promoter in human and mouse. Epigenetics 1: 155-162.

192. Bocchio CL, Boin F, Zanardini R, Popoli M, Michelato A, Bignotti S, Tura GB, and Gennarelli M (2002) Association between promoter polymorphic haplotypes of interleukin-10 gene and schizophrenia. Biol Psychiatry 51: 480-484.

193. Ozbey U, Tug E, and Namli M (2009) Interleukin-10 gene promoter polymorphism in patients with schizophrenia in a region of East Turkey. World J Biol Psychiatry 10: 461-468.

194. Burger GCE, Drummond JC, and Sandstead HR (1948) Malnutrition and Starvation in Western Netherlands, September 1944-July 1945. The Hague: General State Printing Office.

195. Talens RP, Boomsma DI, Tobi EW, Kremer D, Jukema JW, Willemsen G, Putter H, Slagboom PE, and Heijmans BT (2010) Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. FASEB J 24: 3135-3144.

196. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38: 1378-1385.

197. Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris TB, Launer L, Gudnason V, and Fallin MD (2010) Personalized epigenomic signatures that are stable over time and covary with body mass index. Sci Transl Med 2: 49ra67.

198. Bell JT, and Spector TD (2011) A twin approach to unraveling epigenetics. Trends Genet 27: 116-125.

199. Martin N, Boomsma D, and Machin G (1997) A twin-pronged attack on complex traits. Nat Genet 17: 387-392.

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

201. Williamson CM, Turner MD, Ball ST, Nottingham WT, Glenister P, Fray M, et al. (2006) Identification of an imprinting control region affecting the expression of all transcripts in the Gnas cluster. Nat Genet 38: 350-355.

202. Heijmans BT and Mill J (2012) Commentary: The seven plagues of epigenetic epidemiology. Int J Epidemiol 41: 74-78.

203. Purcell S (2002) Variance components models for gene-environment interaction in twin analysis. Twin Res 5: 554-571.

204. Wong CC, Caspi A, Williams B, Craig IW, Houts R, Ambler A, Moffitt TE, and Mill J (2010) A longitudinal study of epigenetic variation in twins. Epigenetics 5: 516-526.

205. Thompson RF, Atzmon G, Gheorghe C, Liang HQ, Lowes C, Greally JM, and Barzilai N (2010) Tissue-specific dysregulation of DNA methylation in aging. Aging Cell 9: 506-518.

206. McEvoy BP, Montgomery GW, McRae AF, Ripatti S, Perola M, et al. (2009) Geographical structure and differential natural selection among North European populations. Genome Res 19: 804-814.

207. Christensen K, Bathum L, and Christiansen L (2008) Biological Indicators and Genetic Information in Danish Twin and Oldest-old Surveys. In: Weinstein M, Vaupel JW, Wachter KW, editors. Biosocial Surveys. Washington: The National Academies Press. pp. 15-41.

208. Skytthe A, Kyvik K, Bathum L, Holm N, Vaupel JW, and Christensen K (2006) The Danish Twin Registry in the new millennium. Twin Res Hum Genet 9: 763-771.

209. Gluckman PD and Hanson MA (2004) Developmental origins of disease paradigm: a mechanistic and evolutionary perspective. Pediatr Res 56: 311-317.

210. Barker DJ (1995) The Wellcome Foundation Lecture, 1994. The fetal origins of adult disease. Proc Biol Sci 262: 37-43.

211. de Rooij S, Painter RC, Roseboom TJ, Phillips DI, Osmond C, Barker DJ, Tanck MW, Michels RP, Bossuyt PM, and Bleker OP (2006) Glucose tolerance at age 58 and the decline of glucose tolerance in comparison with age 50 in people prenatally exposed to the Dutch famine. Diabetologia 49: 637-643.

212. Lillycrop KA, Phillips ES, Jackson AA, Hanson MA, and Burdge GC (2005) Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. J Nutr 135: 1382-1386.

213. Gluckman PD, Lillycrop KA, Vickers MH, Pleasants AB, Phillips ES, Beedle AS, Burdge GC, and Hanson MA (2007) Metabolic plasticity during mammalian development is directionally dependent on early nutritional status. Proc Natl Acad Sci U S A 104: 12796-12800.

214. Rakyan VK, Blewitt ME, Druker R, Preis JI, and Whitelaw E (2002) Metastable epialleles in mammals. Trends Genet 18: 348-351.

215. Einstein F, Thompson RF, Bhagat TD, Fazzari MJ, Verma A, Barzilai N, and Greally JM (2010) Cytosine methylation dysregulation in neonates following intrauterine growth restriction. PLoS ONE 5: e8887.

216. Breton CV, Byun HM, Wenten M, Pan F, Yang A, and Gilliland FD (2009) Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. Am J Respir Crit Care Med 180: 462-467.

217. Thornburg KL, O'Tierney PF, and Louey S (2010) Review: The placenta is a programming agent for cardiovascular disease. Placenta 31 Suppl: S54-S59.

218. Shepherd J, Blauw GJ, Murphy MB, Bollen EL, Buckley BM, Cobbe SM, et al. (2002) Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet 360: 1623-1630.

219. Hoyo C, Murtha AP, Schildkraut JM, Jirtle RL, mark-Wahnefried W, Forman MR, Iversen ES, Kurtzberg J, Overcash F, Huang Z, and Murphy SK (2011) Methylation variation at IGF2 differentially methylated regions and maternal folic acid use before and during pregnancy. Epigenetics 6: 928-936.

220. Kim M, Long TI, Arakawa K, Wang R, Yu MC, and Laird PW (2010) DNA methylation as a biomarker for cardiovascular disease risk. PLoS ONE 5: e9692.

221. Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, Sparrow D, Vokonas P, and Schwartz J (2010) Ischemic heart disease and stroke in relation to blood DNA methylation. Epidemiology 21: 819-828.

222. Barker DJ, Osmond C, Forsen TJ, Kajantie E, and Eriksson JG (2005) Trajectories of growth among children who have coronary events as adults. N Engl J Med 353: 1802-1809.

223. Weinstein LS, Xie T, Zhang QH, and Chen M (2007) Studies of the regulation and function of the G(s) alpha gene Gnas using gene targeting technology. Pharmacol Ther 115: 271-291.

224. Moore GE, bu-Amero SN, Bell G, Wakeling EL, Kingsnorth A, Stanier P, Jauniaux E, and Bennett ST (2001) Evidence that insulin is imprinted in the human yolk sac. Diabetes 50: 199-203.

225. Christian P and Stewart CP (2010) Maternal micronutrient deficiency, fetal development, and the risk of chronic disease. J Nutr 140: 437-445.

226. Power C, Atherton K, and Thomas C (2010) Maternal smoking in pregnancy, adult adiposity and other risk factors for cardiovascular disease. Atherosclerosis 211: 643-648.

227. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, et al. (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet 5: e1000602.

228. Madsen L and Kristiansen K (2010) The importance of dietary modulation of cAMP and insulin signaling in adipose tissue and the development of obesity. Ann N Y Acad Sci 1190: 1-14.

229. Joakimsen O, Bonaa KH, Stensland-Bugge E, and Jacobsen BK (1999) Age and sex differences in the distribution and ultrasound morphology of carotid atherosclerosis: the Tromso Study. Arterioscler Thromb Vasc Biol 19: 3007-3013.

230. Theres H, Maier B, Matteucci GR, Schnippa S, Kallischnigg G, Schuren KP, and Thimme W (2004) Influence of gender on treatment and short-term mortality of patients with acute myocardial infarction in Berlin. Z Kardiol 93: 954-963.

231. Maekawa M and Watanabe Y (2007) Epigenetics: relations to disease and laboratory findings. Curr Med Chem 14: 2642-2653.

232. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 31: 142-147.

233. Talens RP, Jukema JW, Trompet S, Kremer D, Westendorp RG, Lumey LH, Sattar N, Putter H, Slagboom PE, and Heijmans BT (2012) Hypermethylation at loci sensitive to the prenatal environment is associated with increased incidence of myocardial infarction. Int J Epidemiol 41: 106-115.

234. Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S, et al. (2012) Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol 13: R43.

235. Heijmans BT, Westendorp RG, Knook DL, Kluft C, and Slagboom PE (1999) Angiotensin I-converting enzyme and plasminogen activator inhibitor-1 gene variants: risk of mortality and fatal cardiovascular disease in an elderly population-based cohort. J Am Coll Cardiol 34: 1176-1183.

236. Docherty SJ, Davis OS, Haworth CM, Plomin R, and Mill J (2009) Bisulfite-based epityping on pooled genomic DNA provides an accurate estimate of average group DNA methylation. Epigenetics Chromatin 2: 3.

237. Du P, Kibbe WA, and Lin SM (2008) lumi: a pipeline for processing Illumina microarray. Bioinformatics 24: 1547-1548.

238. Wong N, Morley R, Saffery R, and Craig J (2008) Archived Guthrie blood spots as a novel source for quantitative DNA methylation analysis. Biotechniques 45: 423-4, 426, 428.

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

240. Bock C, Halachev K, Buch J, and Lengauer T (2009) EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi)genomic data. Genome Biol 10: R14.

241. Baccarelli A, Tarantini L, Wright RO, Bollati V, Litonjua AA, Zanobetti A, Sparrow D, Vokonas P, and Schwartz J (2010) Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study. Epigenetics 5.

242. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet 6: e1000952.

243. Tobi EW, Slagboom PE, van Dongen J, Kremer D, Stein AD, Putter H, Heijmans BT, and Lumey LH (2012) Prenatal famine and genetic variation are independently and additively associated with DNA methylation at regulatory loci within IGF2/H19. PLoS ONE 7: e37933.

244. Murgatroyd C, Patchev AV, Wu Y, Micale V, Bockmuhl Y, Fischer D, Holsboer F, Wotjak CT, Almeida OF, and Spengler D (2009) Dynamic DNA methylation programs persistent adverse effects of early-life stress. Nat Neurosci 12: 1559-1566.

245. Meulenbelt I, Droog S, Trommelen GJ, Boomsma DI, and Slagboom PE (1995) High-yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. Am J Hum Genet 57: 1252-1254.

246. Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, Smith AV, and Gudnason V (2012) Heterogeneity in white blood

cells has potential to confound DNA methylation measurements. PLoS ONE 7: e46705.

247. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, at al. (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res 20: 434-439.

248. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49: 359-367.

249. Teschendorff AE, West J, and Beck S (2013) Age-associated epigenetic drift: implications, and a case of epigenetic thrift? Hum Mol Genet .

250. Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, and Vilain E (2011) Epigenetic predictor of age. PLoS ONE 6: e14821.

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

252. Zhuang J, Peng W, Li H, Wang W, Wei Y, Li W, and Xu Y (2012) Methylation of p15INK4b and expression of ANRIL on chromosome 9p21 are associated with coronary artery disease. PLoS ONE 7: e47193.

253. Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, Blow MJ, Cohen JC, Rubin EM, and Pennacchio LA (2010) Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. Nature 464: 409-412.

254. Guay SP, Brisson D, Munger J, Lamarche B, Gaudet D, and Bouchard L (2012) ABCA1 gene promoter DNA methylation is associated with HDL particle profile and coronary artery disease in familial hypercholesterolemia. Epigenetics 7: 464-472.

255. Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, et al. (2013) Identification and systematic annotation of tissue-specific differentially

methylated regions using the Illumina 450k array. Epigenetics Chromatin 6: 26.

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

257. Smith GD (2011) Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. Int J Epidemiol 40: 537-562.

258. Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, and Brenner H (2012) Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. Eur Heart J 33: 2841-2848.

259. Kaati G, Bygren LO, and Edvinsson S (2002) Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. Eur J Hum Genet 10: 682-688.

260. Kaati G, Bygren LO, Pembrey M, and Sjostrom M (2007) Transgenerational response to nutrition, early life circumstances and longevity. Eur J Hum Genet 15: 784-790.

261. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, and Turner SW (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7: 461-465.

262. Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, and Clark SJ (2012) Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res 22: 1120-1127.

263. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al., (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet 45: 580-585.

Nederlandse samenvatting

Epigenetische mechanismen reguleren het gebruik van de genen

De chromosomen in een menselijke cel bereiken kop-aan-staart gelegd een lengte van ongeveer 2 meter. Om in een celkern met een diameter van ongeveer 1 micrometer (= 0,001 mm) te passen moet het DNA worden opvouwen. Dit wordt bewerkstelligd door het DNA eerst te wikkelen rondom histon eiwitten. Het koppelen van verschillende kleine chemische groepen aan de histon eiwitten zorgt ervoor dat deze eiwit - DNA complexen elkaar aantrekken of afstoten wat een strakker of losser opgevouwen stuk chromosoom tot gevolg heeft. Ook aan het DNA zelf kan op specifieke plaatsen een chemische groep, een methyl groep, gekoppeld worden wat de affiniteit beïnvloedt tussen DNA en DNA bindende eiwitten zoals transcriptiefactoren. Het geheel van chemische groepen aan en rondom het DNA wordt epigenetische informatie genoemd (Oud Grieks: epi = op).

Een cel maakt geen gebruik van genen die in strak opgevouwen stukken DNA liggen. Simpel gezegd wordt door epigenetische informatie bepaald onder welke omstandigheden de cel bepaalde genen überhaupt kan 'gebruiken' om biologische taken te verrichten. Wanneer we bijvoorbeeld eten wordt insuline alleen gemaakt in de eilandjes van Langerhals van de alvleesklier, terwijl elke cel in ons lichaam hetzelfde gen voor insuline bezit en vele cellen insuline nodig hebben. Tijdens de celdeling wordt zowel de genetische als epigenetische informatie van de oorspronkelijke cel correct doorgegeven aan beide dochter-cellen. Echter, waar het genetisch gereguleerde gen gebruik vanaf de bevruchting vast staat voor elke cel in het lichaam, kan het epigenetisch gereguleerde gen gebruik van een cel veranderen. Een bekend voorbeeld van zulke epigenetische veranderingen is de ontwikkeling van gespecialiseerde cellen uit stamcellen (met een bevruchte eicel als de ultieme stamcel). Tijdens dit proces leiden omvangrijke epigenetische herstructureringen tot het strak opvouwen van verschillende stukken DNA in dochtercellen waardoor genetisch identieke cellen er toch anders uit zien, andere biologische functies verrichten en verschillend kunnen reageren op signalen uit de omgeving. In een gespecialiseerde cel is epigenetische informatie over het algemeen stabiel, maar de mogelijkheid tot veranderingen blijft. Willekeurige

epigenetische veranderingen kunnen spontaan voorkomen of ontstaan als gevolg van een schadelijke blootstelling, enigszins vergelijkbaar met genetische mutaties. Echter, in tegenstelling tot genetische mutaties, kunnen epigenetische veranderingen ook gericht optreden als onderdeel van een intern proces (b.v. wondgenezing) of als aanpassing op een omgevingsfactor (b.v. eetgedrag).

Doordat epigenetische processen betrokken zijn bij de wijze waarop de cel de genen gebruikt, is het ook te verwachten dat epigenetische veranderingen een rol spelen bij het ontstaan van ziekten. Aangetoond is dat de genen die celdeling en groei controleren epigenetisch volledig ontregeld zijn in tumoren, en ook in de hartspier zijn epigenetische verschillen gevonden tussen cellen in gezonde gedeelten en gedeelten die door een hartinfarct zijn aangetast. Of dit soort epigenetische verschillen oorzaak of gevolg zijn van de ziekte is nog onduidelijk. Wel is het zo dat een opeenstapeling van willekeurige en gerichte epigenetische veranderingen gedurende het leven algemeen wordt gezien als een belangrijke factor in het proces van veroudering en het ontstaan van ouderdomsziekten zoals ouderdomsdiabetes (diabetes type 2), hoge bloeddruk, hart- en vaatziekten (HVZ) en kanker.

Epigenetisch onderzoek naar hart- en vaatziekten

Er is veel interesse in onderzoek naar de relatie tussen epigenetische mechanismen en de ontwikkeling van ziekten, zoals ouderdomsziekten, waarbij embryonale omgeving, leefomgeving en levensstijl belangrijke risicofactoren in zijn. Onderzoek hiernaar in mensen geschiedt veelal via populatiestudies, waarvan er verschillende typen zijn. Er zijn talloze klinische studies die zich specifiek richten op onderzoek naar het effect van (nieuwe) medicatie op het risico om binnen enkele jaren de ziekte te ontwikkelen. Ook zijn er grootschalige populatiestudies die de levensloop, inclusief ziekte en gezondheid, van een grote groep individuen of families gedurende langere tijd onderzoeken. Er zijn zelfs enkele historische cohortstudies die zich concentreren op een specifieke gebeurtenis in het verleden, zoals een hongersnood, en de gevolgen daarvan voor de gezondheid gedurende het leven. Voor de meeste humane studies is biologisch materiaal van de deelnemers, vaak DNA uit bloed, goed beveiligd opgeslagen in zogenaamde

biobanken. Door technologische vooruitgang is epigenetisch onderzoek in humane cohorten, epigenetische epidemiologie genoemd, mogelijk geworden.

Op termijn zal epigenetische epidemiologie meer inzicht verschaffen in de manier waarop de epigenetische verschillen tussen mensen bijdragen aan het ontstaan van ouderdomsziekten. Dit proefschrift vormt een eerste stap in het onderzoek naar de rol van epigenetische processen bij de ontwikkeling van hart- en vaatziekten (HVZ), en bevat twee aspecten. Ten eerste bestuderen en beschrijven we biologische eigenschappen van epigenetische informatie voor een set HVZ kandidaat genen, genen waarvan de betrokkenheid bij HVZ bekend is of vermoed wordt. We onderzochten voor deze genen de lange termijn effecten van ondervoeding voor de geboorte op epigenetische informatie (Hoofdstuk 3) en de veranderlijkheid van epigenetische informatie gedurende de volwassen periode (Hoofdstuk 4). Vervolgens onderzochten we of epigenetische informatie met betrekking tot de prentaal gevoelige set genen het risico op hart- en vaatziekte beïnvloedt (Hoofdstuk 5).

Embryonale omgeving beïnvloedt epigenetische informatie

Epigenetische veranderingen als gevolg van of aanpassing aan omgevingsfactoren kunnen gedurende het hele leven voorkomen. Echter, doordat omvangrijke epigenetische herstructureringen cruciaal zijn bij de embryonale ontwikkeling wordt dit proces gezien als een periode waarin epigenetische informatie erg gevoelig is voor omgevingsinvloeden zoals bijvoorbeeld voedings- of rookgedrag van de moeder (zie Figuur 1, Introductie). Dit geldt in feite ook voor de rijping van de geslachtcellen. Verschillende bevolkingsonderzoeken observeerden een relatie tussen slechte omstandigheden vroeg in het leven en het risico op ouderdomsziekten zoals HVZ, diabetes type 2 en hoge bloeddruk op latere leeftijd. Deze observaties leidden tot formulering van de Developmental Origins of Health and Disease (DOHaD) hypothese die stelt dat een deel van het risico op hart- en vaatziekten bepaald wordt door omstandigheden van voor en vlak na de geboorte. Verder heeft onderzoek in proefdieren aangetoond dat blootstelling voor de geboorte aan bijvoorbeeld te weinig voedsel of voedsel van slechte kwaliteit leidt tot subtiele epigenetische veranderingen die samengaan met een hoger risico op deze

ziekten. Dit soort resultaten toont aan dat epigenetische aanpassing een geloofwaardig moleculair mechanisme is voor de DOHaD hypothese. Echter, doordat proefdieren vrij kort leven in vergelijking met mensen is wel nog onduidelijk hoe blijvend dit soort veranderingen zijn over een periode van tientallen jaren.

Eerder onderzoek van ons heeft aangetoond dat de epigenetische informatie voor een HVZ kandidaat gen blijvend veranderd kan zijn bij ouderen die prenataal blootgesteld waren aan de Hongerwinter van 1944-1945. We onderzochten of dit bij andere kandidaat genen ook het geval kon zijn (Hoofdstuk 3). We vonden dat 6 van de 15 onderzochte genen ook gevoelig waren voor ongunstige omstandigheden tijdens het begin van de zwangerschap. Hierbij viel op dat de richting van de epigenetische verandering niet voor elk gen gelijk was, in tegenstelling tot wat je zou verwachten bij een epigenetische beschadiging. Voor sommige genen was de methylatie toegenomen, bij andere afgenomen ten opzichte van broers en zusters die niet prenataal waren blootgesteld aan de ongunstige omstandigheden. Verder was er voor 5 van de 6 genen een andere of geen epigenetische verandering te zien bij blootstelling aan de Hongerwinter tijdens het einde van de zwangerschap. Dit wijst erop dat de timing van de blootstelling medebepalend is voor het al of niet optreden van de verandering. Ook was het zo dat voor 3 van de 6 genen het effect specifiek in één geslacht voorkwam of veel sterker was ten opzichte van het andere geslacht. Al deze observaties wekken de indruk dat er blijvende epigenetische aanpassingen plaatsvinden in het DNA van cellen als gevolg van blootstelling aan omstandigheden voor de geboorte.

Epigenetische veranderingen tijdens het ouder worden

De resultaten uit de Hongerwinter studies waren gevonden in individuen van 60 jaar oud (Hoofdstuk 3). De mogelijkheid tot epigenetische veranderingen blijft echter gedurende het hele leven bestaan, waardoor we ons afvroegen in hoeverre epigenetische instabiliteit over een dergelijke periode doorwerkt op de HVZ kandidaat genen. Willekeurige epigenetische veranderingen, zowel spontane (bijvoorbeeld door een kopieerfoutje) als die ten gevolge van een schadelijke blootstelling (bijvoorbeeld een niet dodelijke vergiftiging), zijn meestal neutraal

van aard. Een ongecorrigeerde opeenstapeling ervan, “epigenetic drift” genoemd, kan op termijn wel leiden tot minder goed functionerende cellen en weefsels. Algemeen wordt aangenomen dat het vermogen om willekeurige epigenetische foutjes te corrigeren tijdens de veroudering vermindert.

We onderzochten in hoeverre epigenetische informatie voor deze HVZ kandidaat genen verandert gedurende de volwassenheid in een groep eeneiige tweelingen van 18 jaar tot 88 jaar oud (Hoofdstuk 4). We vonden een toename van epigenetische verschillen (variatie) met het ouder worden, zowel tussen niet de verwante individuen als binnen de eeneiige tweelingparen, die in principe genetisch identiek zijn. Dit duidt op een opeenstapeling van epigenetische veranderingen gedurende het leven. Of de veranderingen het gevolg zijn van externe omgevingsfactoren zoals beroep en leefstijl of dat het om willekeurige epigenetische veranderingen gaat was niet te achterhalen. Wel viel op dat niet elk kandidaat gen evenveel epigenetische veranderingen onderging, voor sommige genen was de epigenetische variatie niet veel groter in de oudste groep vergeleken met de jongvolwassenen, terwijl voor andere genen die variatie praktisch verdubbeld was. Verder viel op dat er zelfs in de oudste groep nog tweelingparen waren die epigenetisch nauwelijks van elkaar verschilden. Dit alles duidt erop dat epigenetische variatie verandert gedurende het leven, maar dat niet elk gen even sterk verandert en dat sommige individuen nauwelijks veranderen.

Epigenetische informatie is geassocieerd met risico op hartinfarct

In navolging van het onderzoek in proefdieren naar de betrokkenheid van epigenetische mechanismen in de relatie tussen omstandigheden voor de geboorte en het risico op HVZ (DOHaD hypothese) vroegen we ons af of de epigenetische veranderingen als gevolg van de hongerwinter (Hoofdstuk 3) ook betrokken kunnen zijn bij de ontwikkeling van HVZ. Uit verschillende onderzoeken zijn aanwijzingen gekomen dat individuen die voor de geboorte blootgesteld waren aan de Hongerwinter een hoger risico hebben op het ontstaan van HVZ. De Hongerwinter familiestudie is echter door haar unieke karakter (zie Box 1 Introduction) aan de kleine kant voor dit soort epidemiologisch onderzoek. Verder

hebben de blootgestelde individuen nog niet de leeftijd bereikt waarop de meeste mensen voor het eerst serieuze HVZ problemen ervaren, waardoor een degelijke vergelijking van hun epigenetische profiel met dat van niet blootgestelde individuen binnenkort pas een reële mogelijkheid is.

We onderzochten in een klinische studie in hoeverre de epigenetische informatie voor de zes HVZ kandidaat genen die gevoelig bleken voor omstandigheden voor de geboorte (Hoofdstuk 3), betrokken is bij het risico op het krijgen van een hartinfarct op latere leeftijd (Hoofdstuk 5). We vonden dat epigenetische informatie op 2 van deze genen verschilde tussen vrouwen die gedurende de vervolgperiode van drie jaar een hartinfarct kregen en vrouwen met een vergelijkbaar HVZ risico die hiervan gevrijwaard bleven. Voor mannen vonden we geen verschillen. Verder vonden we dat diegenen met het epigenetische verschil op beide genen het hoogste risico hadden op een hartinfarct. Het epigenetische verschil was niet te verklaren door andere HVZ risicofactoren zoals zwaarlijvigheid en de bloedspiegels van het cholesterol. Dit alles duidt erop dat epigenetische informatie een nieuwe indicator voor het risico op hart- en vaatziekten kan zijn, die wellicht deels afhangt van omstandigheden voor de geboorte.

Praktische strategieën voor een nieuw onderzoeksveld

Naast het biologische aspect bevat dit proefschrift ook een tweede, meer methodologisch aspect, waarin we praktische strategieën beschrijven voor het uitvoeren van epigenetisch onderzoek op humane cohortstudies. Dit nieuwe onderzoeksveld wordt epigenetische epidemiologie genoemd en is volop in ontwikkeling. Recent is er een aantal projecten gelanceerd om biobanken die specifiek ontworpen zijn voor epigenetisch onderzoek op te zetten met materiaal uit meerdere toegankelijke weefsels zoals bloed, wangslijmvlies, en opperhuid. Het opzetten van biobanken kost echter veel tijd en middelen, waardoor epigenetisch epidemiologische studies vooralsnog zijn aangewezen op het materiaal van de bestaande biobanken die veelal ontworpen zijn voor genetisch onderzoek. Ondanks dat zulk epigenetisch onderzoek technisch uitvoerbaar is, zijn er een aantal praktische complicaties die samenhangen met de dynamische eigenschappen van epigenetische informatie. Waar een genetische bepaling informatie geeft over de genetische code in elke cel van het lichaam gedurende

het hele leven, is dit voor een epigenetische bepaling niet vanzelfsprekend. Ook is het aantal toegankelijke weefsels van levende mensen voor epidemiologisch onderzoek beperkt, en hebben die toegankelijke weefsels (zoals bloed en wangslijmvlies) vaak een beperkte rol bij het ontstaan van ziekte. Dit houdt in dat onduidelijk is welke epigenetische onderzoeksvragen beantwoord kunnen worden middels epidemiologisch onderzoek en voor welke vragen celculturen en diermodellen noodzakelijk zijn.

Tijdens de totstandkoming van de huidige biobanken is bij het verzamelen van het genetisch materiaal geen rekening gehouden met een mogelijke toepassing in epigenetisch onderzoek. We onderzochten in hoeverre en onder welke voorwaarden de enorme collecties biologisch materiaal in deze biobanken, meestal DNA uit bloed, geschikt is voor epigenetisch onderzoek (Hoofdstuk 2). Hiervoor beschreven we eerst de normale variatie en patronen van epigenetische informatie voor 16 HVZ kandidaat genen aan de hand van DNA uit bloed van 30 gezonde individuen tussen 21 en 73 jaar oud (Hoofdstuk 2). Daarna vonden we dat een substantieel deel van deze kandidaat genen voldoet aan de voorwaarden voor gebruik in epigenetische epidemiologie (Hoofdstuk 2). Op geschikte genen is de variatie in epigenetische informatie tussen personen stabiel gedurende de vervolgperiode van een studie (Hoofdstuk 2). Verder wordt de epigenetische variatie niet beïnvloed door persoonlijke verschillen in de samenstelling van het bloed in de vele typen (witte) bloedcellen (Hoofdstukken 2 en 4). Genen die aan deze criteria voldoen kunnen gebruikt worden als indicatoren voor een schadelijke blootstelling en het risico op ziekten (Hoofdstukken 3 en 5).

Het koppelen van een epigenetisch mechanisme aan de ontwikkeling van een ziekte vereist ook kennis over epigenetische informatie in het (interne) weefsel waarin de ziekte ontstaat. Omdat bloed als weefsel niet direct betrokken is bij het ontstaan van veel ziekten dient de epigenetische informatie in bloed wel representatief te zijn voor het aan ziekte gerelateerde weefsel. In een verkennende genoomwijde studie op autopsiemateriaal (bloed, skelet spier, onderhuids vet, buikholte vet, lever, nier, alvleesklier en milt) vonden we dat epigenetische informatie in bloed gedeeltelijk representatief kan zijn voor deze interne weefsels (Hoofdstuk 6). In de meeste gevallen leek de overeenkomst

in epigenetische variatie afhankelijk te zijn van de combinatie van meetpunt en weefsels. Grondige genoomwijde herhaling van al deze experimenten op grotere populaties met monsters van meerdere toegankelijke weefsels zal op termijn van alle kandidaat genen de geschiktheid voor epigenetisch epidemiologisch onderzoek aantonen.

De toekomst van epigenetische epidemiologie

In dit proefschrift hebben we een eerste verkennende stap gezet in het epidemiologische onderzoek naar de rol van epigenetische processen bij de ontwikkeling van HVZ. We vonden dat het dynamische aspect van epigenetische informatie wel beperkingen oplegt aan de praktische uitvoering en interpretatie van dit type onderzoek, maar het niet onmogelijk maakt (Hoofdstukken 2 – 6). Zo vonden we kandidaat genen waarvan de epigenetische variatie in twee weefsels stabiel is over een periode van 20 jaar (Hoofdstuk 2). De opvolgperiode van de meeste klinische studies is veel korter dan dit, wat inhoudt dat deze kandidaat genen als epigenetische indicator kunnen dienen voor het risico op ziekten (Hoofdstuk 5). Verder vonden we dat blootstelling aan de extreme omstandigheden van de Hongerwinter tijdens de embryonale ontwikkeling blijkt te leiden tot een permanente beschadiging of aanpassing van epigenetische informatie op sommige kandidaat genen (Hoofdstuk 3). In een tweede studie vonden we dat epigenetische informatie op dit soort kandidaat genen ook een indicator kan zijn van het risico op HVZ (Hoofdstuk 5), een ziekte waarvan het ontstaan dikwijls in verband gebracht is met omstandigheden vroeg in het leven (DOHaD hypothese). Ook voor minder extreme omstandigheden tijdens de zwangerschap, zoals bijvoorbeeld roken of verminderde functie van de placenta, is gevonden dat epigenetische veranderingen een gevolg kunnen zijn. In dit proefschrift vonden we dat juist epigenetisch onderzoek geschikt blijkt om via een combinatie van verschillende studies in kaart te brengen wat de uiteindelijke gevolgen van deze omstandigheden op latere leeftijd zijn (Hoofdstukken 3 – 5).

De resultaten in dit proefschrift kunnen ook van waarde zijn voor het aanleggen van nieuwe biobanken of het uitbreiden van bestaande biobanken om er meer epigenetische vragen mee te kunnen beantwoorden. Het is voor grotere populatiestudies onhaalbaar om elk

celtype van een weefsel apart te meten, en dat hoeft voor de meeste kandidaat genen ook niet (Hoofdstuk 2). Wel is het van belang dat er per individu gegevens zijn over de aantallen van de belangrijkere celtypen in bloed, om rekening te kunnen houden met de invloed van onderlinge verschillen in de cellulaire compositie die aan epigenetische verschillen tussen mensen ten grondslag kunnen liggen (Hoofdstukken 2 en 4). Verder geven onze resultaten aan dat via het verzamelen van meerdere toegankelijke weefselmonsters meer epigenetische informatie te ontrafelen is over ontoegankelijke, maar ziekte gerelateerde weefsels (Hoofdstuk 6). Bij het verzamelen van informatie over de deelnemers lijkt het waardevol om informatie over dieet en leefomstandigheden in de vroege kindertijd (Hoofdstuk 3), aan het beroep gerelateerde omgevingsfactoren, of aan de gekozen leefstijl te vragen. Verder komt uit dit proefschrift sterk naar voren dat er behoefte is aan databases waarmee de genoomwijde epigenetische variatie tussen personen, tussen de weefsels binnen een persoon en de veranderlijkheid van deze variatie binnen een weefsel over de tijd geraadpleegd kan worden. Er wordt op dit moment veel geïnvesteerd in het creëren van zulke databases waardoor het te verwachten is dat dit soort informatie in de nabije toekomst steeds beter voorhanden is.

Desalniettemin zal niet elke epigenetische vraag middels humane populatiestudies beantwoord kunnen worden. Het zou ook teveel zijn dit te verwachten. Dit proefschrift toont echter aan dat met aanpassing van bekende (klassieke) epidemiologische studieontwerpen en het zorgvuldig combineren van populatiestudies er epigenetisch onderzoek mogelijk is waarin omstandigheden en ziekten over een heel leven aan elkaar kunnen worden gekoppeld. Als daarbij ook de kennis van genetische erfelijkheid en genexpressie geïncorporeerd kan worden, kunnen een aantal belangrijke vragen over het ontstaan van verouderingsziekten systematisch worden onderzocht. Samenvattend belooft de epigenetische epidemiologie in de nabije toekomst een boeiende en betekenisvolle bijdrage te leveren aan het ontrafelen van de complexe interactie tussen gen en omgeving die uiteindelijk de sleutel vormt voor het ontstaan van complexe verouderingsziekten zoals HVZ.

Dankwoord

In feite is promoveren een eenzame bezigheid die voor de buitenwereld obsessief kan lijken en, nu ik zo terug kijk, dat af en toe vermoedelijk ook wel was. Ik heb veel plezier beleefd aan het onderzoek en alles wat daarbij kwam kijken, maar het vergde veel van mij en ik hield soms weinig ruimte over voor iets anders. Dat mijn leven zoveel leuke facetten en ontwikkelingen kende heb ik volledig te danken aan de steun, ruimte en betrokkenheid die ik de afgelopen jaren van mijn gezin, familie, vrienden en collega's heb gekregen. Aan het eind van dit proefschrift wil ik hiervoor dan ook iedereen bedanken. Iedereen persoonlijk bedanken zou veel ruimte innemen en ik vind dit proefschrift wel lang genoeg. Ik hoop dan ook dat je je hiermee toch aangesproken voelt als ik je in de verdere tekst niet bij naam noem. **Bedankt** voor alle lol, alle interesse, alle steun en alle begrip.

Een aantal mensen wil ik hier graag persoonlijk bedanken. Laat ik beginnen met mijn promotoren.
Geachte Prof. Dr. Slagboom, beste Eline, wat kun jij toch steeds snel de vinger op de pijnlijke punten van een op het oog goed idee leggen! Niet altijd leuk, maar heel leerzaam. Ik wil je vooral bedanken voor de creatieve vrijheid die ik kreeg en de raad en daad waarmee je me hielp er wetenschap mee te maken. Geachte Prof. Dr. Jukema, beste Wouter, bedankt voor je nooit aflatende enthousiasme voor mijn onderzoeksplannen en het inzicht in de medische wetenschap.
Geachte Dr. Heijmans, beste Bas, ik vond het geweldig om samen met jou en Elmar uit te zoeken hoe epigenetisch epidemiologisch onderzoek is uit te voeren. Ik ben trots op wat we bereikt hebben, bedankt voor alle stimulerende gesprekken.
Om meteen verder te gaan met mijn verdere collega's van Moleculaire Epidemiologie. Dennis, Eka, Ruud, Wesley, Nico, Anika, Joris, Peter, Erik, Yolanda, Steffan, Rita, Willemijn, Yanju, Kai, Eric-Wubbo, Marian, Ingrid en Inge bedankt voor de gezelligheid, jullie enthousiasme, jullie uitstekende hulp in het lab, achter de computer, allerlei regeldingetjes en natuurlijk voor de onvervalste MolEpi sound (close disharmony at its best and worst).
Maurits, het was me een waar genoegen om jou met je eindstage te begeleiden. Ik heb er denk ik net zo veel van geleerd als jij.

Verder wil ik van de afdelingen Biostatistiek en Bioinformatica, Cardiologie en Pathologie binnen het LUMC de volgende mensen bedanken: Geachte Prof. Dr. Putter, beste Hein, bedankt voor de statistische therapie, ik voelde me er na onze gesprekken altijd weer opgelucht over dat ik blijkbaar nog steeds wist waar ik mee bezig was. Linear mixed models for everything! Jelle, ik kan niet wachten tot je het woordenboek "Programmers Gibberish to English" uitbrengt, bedankt voor al je vertalingen en advies. Stella, bedankt voor je enthousiasme. Veel dank aan de anatomen, R. Guijt, T. Alberts en E. Dreef, voor de leerzame lessen anatomie en het bevredigen van mijn nieuwsgierigheid aan de obductietafels. I would further extend my gratitude to Prof. Dr. Boomsma en Dr. Willemsen of the Netherlands Twin Register and the department of biological psychology at the VU university in Amsterdam and Prof. Dr. Christensen and Prof. Dr. Christiansen of the Danish ageing research center and the Danish Twin Registry at the University of Southern Denmark. Dear Dorret, Gonneke, Kaare and Lene thank you for your ideas, your advice, your enthusiasm, and for access to your unique samples and datasets that enabled the epigenetic twin research of chapters 2 and 4 in this thesis.

Besides science, I was also involved in a few other activities at the LUMC that would not have been such a success nor such a lot of fun were it not for the help of the following individuals. Sjoerd, Dwi, Mark, Marion, Elena, Tommy, Boet, and Madeleine, I really enjoyed organizing the 2009 MGC Workshop in Bruges, Belgium with you guys. Godelieve, Cheryl, Dimitris, Petra and Dwi, I can say the same for running the monthly Division 5 PhD meetings with you guys, I like to think that one activity followed from the other.

Dit brengt mij bij vrienden en familie. Jochen, we kennen elkaar al vanaf onze kinderjaren en alweer ben je betrokken bij een belangrijke gebeurtenis in mijn leven. Antoine, ik mis onze gesprekken sinds je in het verre Palo Alto zit. Rinke, bedankt voor het druk-klaar maken van dit proefschrift en voor het ontwerpen van de voorkant. Saskia, bedankt voor het lenen van Rinke. Joost (Fiepke), Eva, Gijs, Nelleke, Daan, Joost, Annelies, Dennis, Suzanne, Annefloor, Jeroen, Tess, Pelin, en alle anderen, bedankt voor alle gezelligheid waarmee jullie mij de noodzakelijke

afleiding gaven. Pa, ma, Frank, Geraldien (Schoonmams), Frits (Schoonpa), Charlotte en Edwin, ik vond het erg fijn dat jullie bleven luisteren naar iets dat voor mij tijdens de hele opbouwfase niet altijd even helder was. Hopelijk geeft dit boekje jullie ook iets meer houvast.

En dan nu de laatste en twee belangrijkste personen in mijn leven. Simone, jouw liefde, steun, trots, en aanmoedigingen hielpen me om steeds weer achter de computer te kruipen en verder te werken als iets anders doen eigenlijk veel aantrekkelijker was (weer een zonnige zondag op de zolderkamer, zucht). Ik ben dankbaar voor hoe je met me meeleefde. Je luisterde overduidelijk naar mijn enthousiasme in dit werk, want ik kon aan je ogen zien dat je de inhoud vaak allang niet meer volgde.
Dit proefschrift heeft evenveel van jou als van mij gevergd. Lieve Rolf, mijn kleine, grote, stoere bink, ik geniet van jouw energieke ontdekkingstocht door de wereld. Je bent gek op puzzelen en uit eigen ervaring weet ik nu hoe nuttig dat soms kan zijn.

Curriculum Vitae

Rudolf Pieter Talens was born on the 11th of October in Eindhoven, The Netherlands. In 1995 he received his grammar school diploma at 'Gymnasium Augustinianum' in Eindhoven and started his study of Biology at the Radboud University Nijmegen, The Netherlands. After a pause due to private circumstances, he resumed his studies in 2002, specializing in molecular biology. His first research internship was at the department of Microbiology at the Radboud University which introduced him to genetic research. Here he investigated the emergence of the eukaryote through comparing sequences of mitochondrial genes of an anaerobic eukaryote with homologous genes in alpha-proteobacteria. His main research internship was at the department of plant genetics where he developed a method to track loss of heterozygosity events at the genotype level in petunia. At his own initiative both internships were extended with several months and he graduated in 2005. In 2006 he worked as a research associate at the Université de Montréal, Canada. In 2007 he started his PhD research at the department of of molecular epidemiology under the supervision of Prof. slagboom, Prof Jukema, and Dr. Heijmans. The project, funded by the Dutch Heart Foundation, aimed to investigate the relation between epigenetic variation and Cardiovascular disease. The results of this research are outlined in this Thesis. From 2012 to 2014 he developed and taught several courses on molecular biology to international undergraduate students at the HZ univeristy of applied sciences in Vlissingen, The Netherlands. Here he also initiated a researchline on the use of biomolecular techniques in on shore cultivation of mollusca. From 2013 to 2014 he introduced high school teachers to the field of epigenetic research, in cooperation with the Beta Plaza Zeeland. He is now working towards a career shift into the field of data analysis.

Publications

Roderick C Slieker, Steffan D Bos, Jelle J Goeman, Judith VMG Bovée, Rudolf P Talens, Ruud van der Breggen, H Eka D Suchiman, Eric-Wubbo Lameijer, Hein Putter, Erik B van den Akker, Yanju Zhang, J Wouter Jukema, P Eline Slagboom, Ingrid Meulenbelt, and Bastiaan T Heijmans (2013) *Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array*. **Epigenetics & Chromatin**; *Vol 6, Iss 1: article 26; Pg 1-12*

Rudolf P. Talens, Kaare Christensen, Hein Putter, Gonneke Willemsen, Lene Christiansen, Dennis Kremer, H. Eka D. Suchiman, P. Eline Slagboom, Dorret I. Boomsma, and Bastiaan T. Heijmans (2012) *Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs*. **Aging Cell**; *Vol 11, Iss 4; Pg 694 – 703*

RP Talens, JW Jukema, S Trompet, D Kremer, RGJ Westendorp, LH Lumey, N Sattar, H Putter, PE Slagboom, and BT Heijmans on behalf of the PROSPER group (2012) *Hypermethylation at loci sensitive to the prenatal environment is associated with increased incidence of myocardial infarction in women*. **International Journal of Epidemiology**; *Vol 41, Iss 1; Pg 106 – 115*

M. Lourdes Sampietro, Stella Trompet, Jeffrey J.W. Verschuren, Rudolf P. Talens, Joris Deelen, Bastiaan T. Heijmans, Robbert J. de Winter, Rene A. Tio, Pieter A.F.M. Doevendans, Santhi K. Ganesh, Elizabeth G. Nabel, Harm-Jan Westra, Lude Franke, Erik B. van den Akker, Rudi G.J. Westendorp, Aeilko H. Zwinderman, Adnan Kastrati, Werner Koch, P.Eline Slagboom4, Peter de Knijff, and J. Wouter Jukema (2011) *A genome-wide association study identifies a region at chromosome 12 as a potential susceptibility locus for restenosis after Percutaneous Coronary Intervention*. **Human Molecular Genetics**; *Vol 20, Iss 2;, Pg 4748 – 4757*

Rudolf P. Talens, Dorret I. Boomsma, Elmar W. Tobi, Dennis Kremer, J. Wouter Jukema, Gonneke Willemsen, Hein Putter, P. Eline Slagboom, and Bastiaan T. Heijmans (2010) *Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology*. **FASEB Journal**; *Vol 24, Iss 9; Pg 3135 – 3144*

Elmar W. Tobi, L.H. Lumey, Rudolf P. Talens, Dennis Kremer, Hein Putter, Aryeh D. Stein, P. Eline Slagboom, and Bastiaan T. Heijmans (2009) *DNA Methylation differences after exposure to prenatal famine are common and timing- and sex-specific*. **Human Molecular Genetics**; *Vol 18, Iss 2; Pg 4046 – 4053*