



Universiteit
Leiden
The Netherlands

Studies into epigenetic variation and its contribution to cardiovascular disease

Talens, R.P.

Citation

Talens, R. P. (2015, January 8). *Studies into epigenetic variation and its contribution to cardiovascular disease*. Retrieved from <https://hdl.handle.net/1887/30776>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/30776>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/30776> holds various files of this Leiden University dissertation

Author: Talens, Rudolf P.

Title: Studies into epigenetic variation and its contribution to cardiovascular disease

Issue Date: 2015-01-08

Chapter 2

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

Rudolf P. Talens¹, Dorret I. Boomsma², Elmar W. Tobi¹, Dennis Kremer¹, J. Wouter Jukema³, Gonneke Willemsen², Hein Putter⁴, P. Eline Slagboom¹, 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

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 dinucleotides) 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, representative 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 ≥ 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 = $\text{Arc tan} \left(\frac{\text{Methylation}}{(1-\text{Methylation})^2} \right)$. 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
<i>IL10^a</i>	01q32.1	205.01	Anti-inflammation	4	2			+				3	6	+	+ [136]
<i>NR3C1</i>	05q31.3	142.67	Stress response	20	4			+	+			3	2		+ [93]
<i>TNF</i>	06p21.33	31.65	Pro-inflammatory	8	5				+			2			+ [138]
<i>IGF2R^a</i>	06q25.3	160.35	Growth/Apoptosis	10	0	?		+	+	+		1			+ [139]
<i>GRB10</i>	07p12.2	50.82	IIS inhibitor	16	5	+	+	+				4	2		+ [140]
<i>LEP^a</i>	07q32.1	127.67	Metabolism	10	5			+	+				3		+ [141]
<i>CRH^a</i>	08q13.1	67.25	Stress response	5	5				+			1	1		+ [142]
<i>ABCA1</i>	09q31.1	106.73	Cholesterol transport	16	3			+	+			4	4		? [143]
<i>IGF2^a</i>	11p15.5	2.13	Early growth	5	3	+			+	+					+ [133]
<i>INSIGF^a</i>	11p15.5	2.14	(Embryonic) Growth	4	4	?			+			2			+ [144]
<i>KCNQ1OT1^a</i>	11p15.5	2.68	Imprinting control region	14	7	+	+	+							+ [145]
<i>MEG3</i>	14q32.2	100.36	Growth suppressor	7	3	+	+	+				1	1		+ [146]
<i>FTO</i>	16q12.2	52.38	Development	10	4					+				+	
<i>APOC1^a</i>	19q13.32	50.11	Metabolism	6	6						+	3			? [147]
<i>GNASAS</i>	20q13.32	56.86	Growth/Lypolytic signal	17	3				+				1		+ [148]
<i>GNAS A/B</i>	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

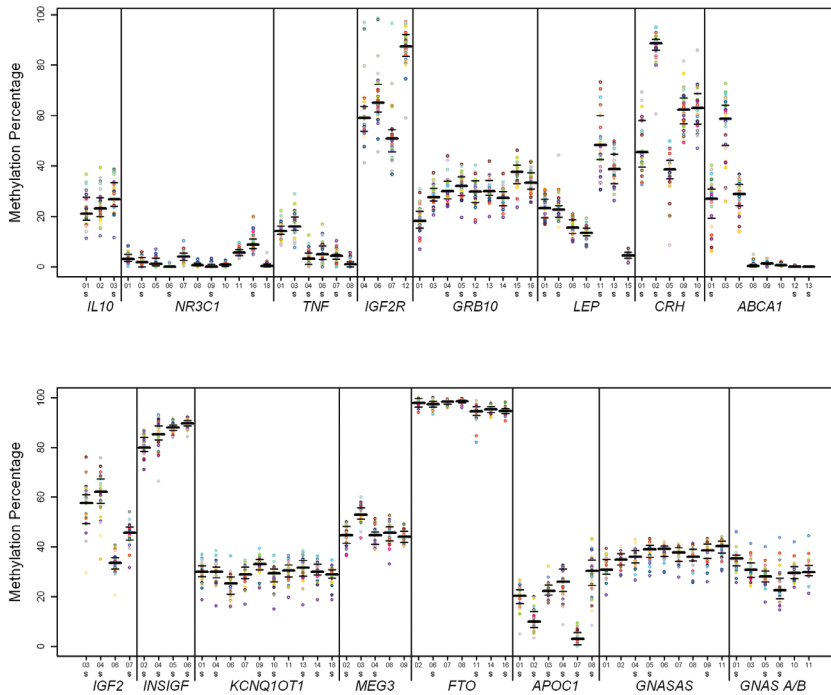


Figure 2.1: Inter-individual variation in DNA methylation

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
<i>IL10</i>	50.1 %	3.9*10 ⁻⁰⁶
<i>NR3C1</i>	0.2 %	0.555
<i>TNF</i>	8.0 %	0.037
<i>IGF2R</i>	5.0 %	0.208
<i>GRB10</i>	0.7 %	0.625
<i>LEP</i>	7.4 %	0.019
<i>CRH</i>	4.0 %	0.022
<i>ABCA1</i>	7.3 %	0.021
<i>IGF2</i>	3.5 %	0.185
<i>INSIGF</i>	0.3 %	0.674
<i>KCNQ1OT1</i>	0.4 %	0.733
<i>MEG3</i>	4.1 %	0.165
<i>FTO</i>	0.1 %	0.714
<i>APOC1</i>	6.2 %	0.026
<i>GNASAS</i>	0.8 %	0.590
<i>GNAS A/B</i>	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*, *INSIGF* 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 ($p=1.0 \times 10^{-4}$). Again, a substantial proportion of the variation in *IL10* methylation could be attributed to the neutrophil percentage (27.9%, $p=8.0 \times 10^{-8}$).

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).

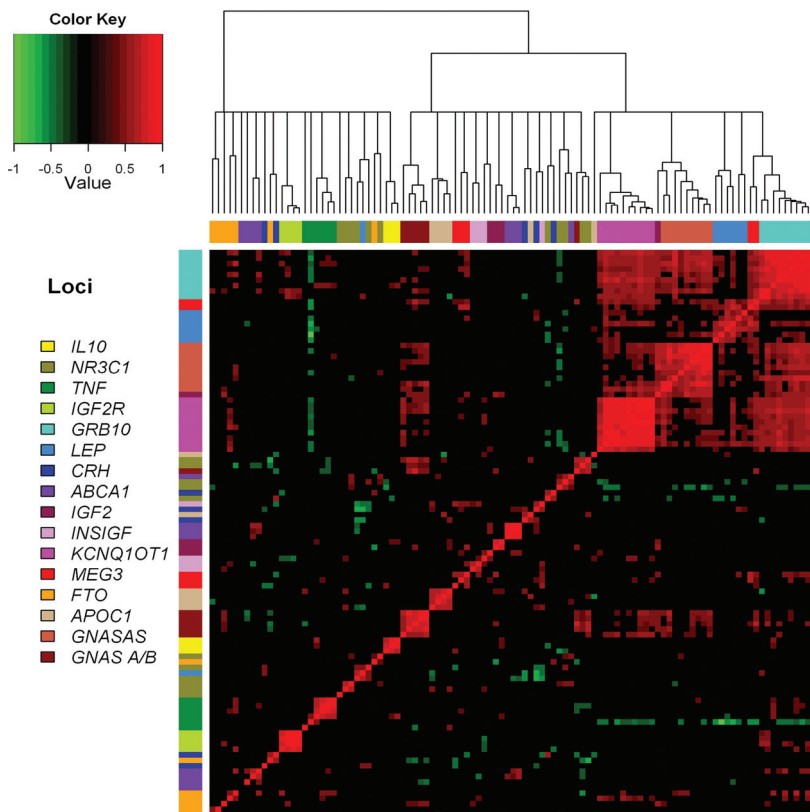


Figure 2.2: Correlation between CpG sites within and across loci

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)

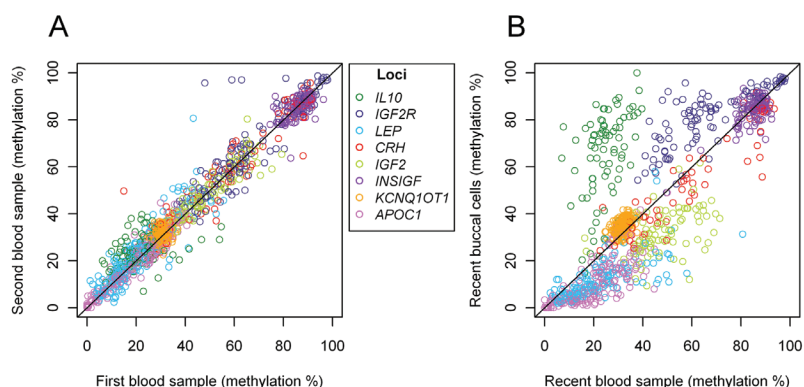


Figure 2.3: Temporal stability and comparison between blood and buccal cell DNA methylation

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

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

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

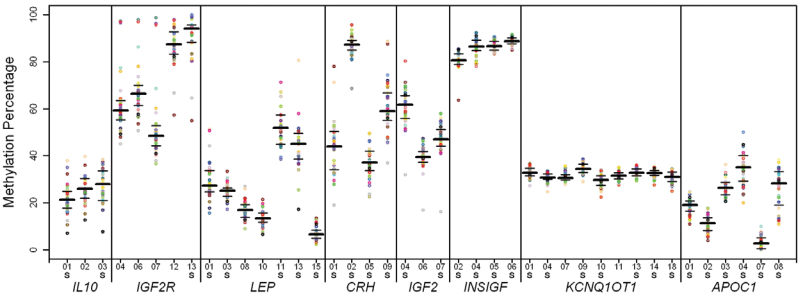


Figure S2.1: Methylation percentage, y-axis, at every CpG-unit, x-axis, for each individual, coloured dots. Measurement was done on DNA from the recent blood samples of the 34 individuals used in the experiments on temporal stability. The order of the loci is based on their chromosomal location, starting 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

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

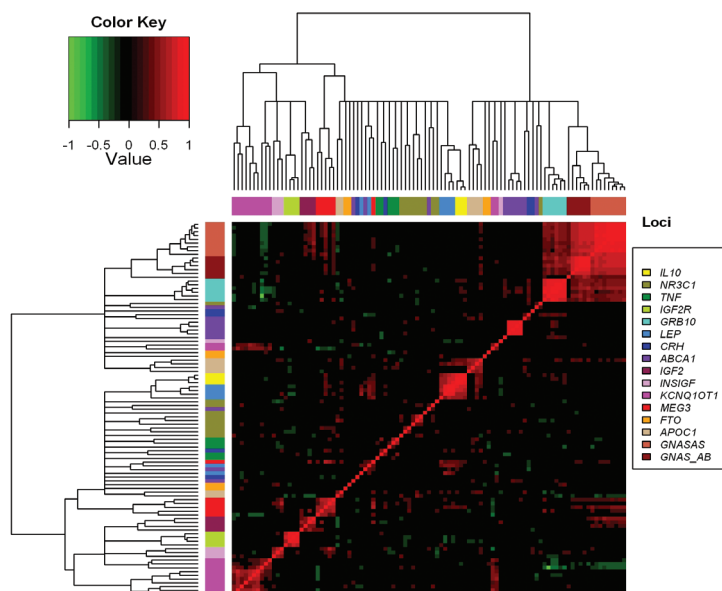


Figure S2.2: Heatmap depicting correlations between the methylation levels of all CpG-units of the 16 loci measured in the controls of the periconceptional group in the Dutch Hunger Winter Families study [14,135] (P.I: Dr. L.H. Lumey, Department of Epidemiology, Mailman School of Public Health, Columbia 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
<i>IL10</i>	TGATTGGTTGAATAT- GAATTTTGTAT	CACCCCCTCATTTTACTTAAAAA
<i>NR3C1</i>	GATTTGGTTTTTTTGGGG	TCCCTTCCCTAAACCT
<i>TNF</i>	GGGTATTTTGTATGTTGT- GTGTT	CAATACTCATAATATCCTTTCCAAAAAA
<i>IGF2R</i>	AGGTAGAAAAAGGTTTGT- GAAG	CAAATCTTAAAACTAACTAAAAACC
<i>GRB10</i>	GGAATTTTAGGATTAAATT- TATGTGA	AACTTCCAAAAAAAACCTCTCC
<i>LEP</i>	GTTTTTGGAGGGATATTA- AGGATT	CTACCAAAAAAACCAACAAAAAA
<i>CRH</i>	TGGTTGTTGTTTTTTGG- TAGG	AATTTCTCCACTCCAAACCTAAA
<i>ABCA1</i>	ATTTTATTGGTGTTTTTG- GTTGT	ATCAAAACCTATACTCTCCCTCCTC
<i>IGF2</i>	TGGATAGGAGATTGAG- GAGAAA	AAACCCCAACAAAAACCACT
<i>INSIGF</i>	GTTTTGAGGAAGAGGT- GTTGA	ACCTAAAATCCAACCACCTAA
<i>KCNQ1OT1</i>	TTTGGTAGGATTTTGTGAG- GAGTTTT	CTCACACCCAACCAATACCTCATAC
<i>MEG3</i>	TTTTTTTAATAGTATTTT- GATTTTTG	AAATAATCCCCACACACATACC
<i>FTO</i>	GTTTGTAAATTTAG- TATTTGGGAGGT	TTTATTTCCATTATCCATTCTCAAA
<i>APOC1</i>	GGAGGAGGGAGATTAATAT- TAATTTGT	ACCCCAAACCTATAACCACTT
<i>GNASAS</i>	GTAATTTGTGGTATGAG- GAAGAGTGA	TAAATAACCCAACTAAATCCCAACA
<i>GNAS A/B</i>	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
<i>IL10</i>	1, 2&3, 4
<i>NR3C1</i>	1&2, 4, 7&8, 9, 10&11, 12&13, 14, 15&16, 17-20, 31, 33&34
<i>TNF</i>	1-3, 5, 6, 9, 10, 11
<i>IGF2R</i>	4&5, 8-10, 11-13, 20&21
<i>GRB10</i>	1&2, 4-6, 7, 8, 17, 18-21, 22&23, 24, 25
<i>LEP</i>	1, 8, 16&17, 19-21, 22, 25, 27
<i>CRH</i>	1, 2, 5, 9, 10
<i>ABCA1</i>	1, 3&4, 6-9, 15&16, 17&18, 19-21, 24, 25
<i>IGF2</i>	3, 4, 6&7, 8
<i>INSIGF</i>	2, 4, 5, 6
<i>KCNQ1OT1</i>	1, 6, 8&9, 10-12, 15, 16, 17&18, 20, 21, 25
<i>MEG3</i>	2, 3, 4, 8&9, 10&11
<i>FTO</i>	2&3, 7, 8&9, 10&11, 14, 17, 19
<i>APOC1</i>	1, 2, 3, 4, 10, 11
<i>GNASAS</i>	1&2, 3&4, 6, 7, 8&9, 10-12, 13&14, 15, 17-19
<i>GNAS A/B</i>	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
<i>IL10</i>	1, 2&3, 4
<i>IGF2R</i>	4&5, 8-10, 11-13, 20&21, 22 ^b
<i>LEP</i>	1, 8, 16&17, 19-21, 22 ^b , 25, 27
<i>CRH</i>	1 ^b , 2, 5, 9
<i>IGF2</i>	4, 6&7, 8
<i>INSIGF</i>	2, 4, 5, 6
<i>KCNQ1OT1</i>	1, 6, 10-12, 15, 16, 17&18, 20, 21, 25
<i>APOC1</i>	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
<i>IL10</i> _01	CpG_1	
<i>IL10</i> _02	CpG_2&3	
<i>IL10</i> _03	CpG_4	
<i>NR3C1</i> _01	CpG_1&2	
<i>NR3C1</i> _02	CpG_3	
<i>NR3C1</i> _03	CpG_4	
<i>NR3C1</i> _04	CpG_5&6	
<i>NR3C1</i> _05	CpG_7&8	
<i>NR3C1</i> _06	CpG_9	
<i>NR3C1</i> _07	CpG_10&11	
<i>NR3C1</i> _08	CpG_12&13	
<i>NR3C1</i> _09	CpG_14	
<i>NR3C1</i> _10	CpG_15&16	
<i>NR3C1</i> _11	CpG_17-20	
<i>NR3C1</i> _12	CpG_21	Mass overlap with unit 14
<i>NR3C1</i> _13	CpG_22-28	
<i>NR3C1</i> _14	CpG_29	Mass overlap with unit 12
<i>NR3C1</i> _15	CpG_30	
<i>NR3C1</i> _16	CpG_31	
<i>NR3C1</i> _17	CpG_32	
<i>NR3C1</i> _18	CpG_33&34	
<i>NR3C1</i> _19	CpG_35-41	High Mass
<i>NR3C1</i> _20	CpG_42	rs5871844 and rs34027900
<i>TNF</i> _01	CpG_1-3	
<i>TNF</i> _02	CpG_4	
<i>TNF</i> _03	CpG_5	
<i>TNF</i> _04	CpG_6	
<i>TNF</i> _05	CpG_7&8	
<i>TNF</i> _06	CpG_9	
<i>TNF</i> _07	CpG_10	
<i>TNF</i> _08	CpG_11	
<i>IGF2R</i> _01	CpG_1	Mass overlap with unit 8
<i>IGF2R</i> _02	CpG_2	Mass overlap with unit 3
<i>IGF2R</i> _03	CpG_3	Mass overlap with unit 2
<i>IGF2R</i> _04	CpG_4&5	

Locus_CpG-unit^a	CpG-sites^a	Reason for removal prior to quality control
<i>IGF2R_05</i>	CpG_6&7	Mass overlap with unit 11
<i>IGF2R_06</i>	CpG_8-10	
<i>IGF2R_07</i>	CpG_11-13	
<i>IGF2R_08</i>	CpG_14	Mass overlap with unit 1
<i>IGF2R_09</i>	CpG_15&16	Mass overlap with unit 10 and rs677882 and rs8191722
<i>IGF2R_10</i>	CpG_17	Mass overlap with unit 9
<i>IGF2R_11</i>	CpG_18&19	Mass overlap with unit 5 and rs8191721 and rs8191720
<i>IGF2R_12</i>	CpG_20&21	
<i>IGF2R_13</i>	CpG_22	
<i>GRB10_01</i>	CpG_1&2	
<i>GRB10_02</i>	CpG_3	Mass overlap with unit 9
<i>GRB10_03</i>	CpG_4-6	
<i>GRB10_04</i>	CpG_7	
<i>GRB10_05</i>	CpG_8	
<i>GRB10_06</i>	CpG_9&10	Mass overlap with unit 8
<i>GRB10_07</i>	CpG_11	Low mass
<i>GRB10_08</i>	CpG_12	Mass overlap with unit 6
<i>GRB10_09</i>	CpG_13	Mass overlap with unit 2
<i>GRB10_10</i>	CpG_14&15	
<i>GRB10_11</i>	CpG_16	
<i>GRB10_12</i>	CpG_17	
<i>GRB10_13</i>	CpG_18-21	
<i>GRB10_14</i>	CpG_22&23	
<i>GRB10_15</i>	CpG_24	
<i>GRB10_16</i>	CpG_25	
<i>LEP_01</i>	CpG_1	
<i>LEP_02</i>	CpG_2-7	High Mass and rs791620
<i>LEP_03</i>	CpG_8	
<i>LEP_04</i>	CpG_9&10	Mass overlap with unit 9
<i>LEP_05</i>	CpG_11	Mass overlap with units 12 and 6
<i>LEP_06</i>	CpG_12&13	Mass overlap with units 5 and 12
<i>LEP_07</i>	CpG_14&15	Mass overlap with unit 14
<i>LEP_08</i>	CpG_16&17	
<i>LEP_09</i>	CpG_18	Mass overlap with unit 4

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

Locus_CpG-unit^a	CpG-sites^a	Reason for removal prior to quality control
<i>IGF2_03</i>	CpG_3	rs4930041
<i>IGF2_04</i>	CpG_4	
<i>IGF2_05</i>	CpG_5	
<i>IGF2_06</i>	CpG_6&7	
<i>IGF2_07</i>	CpG_8	
<i>INSIGF_01</i>	CpG_1	Low mass
<i>INSIGF_02</i>	CpG_2	
<i>INSIGF_03</i>	CpG_3	
<i>INSIGF_04</i>	CpG_4	
<i>INSIGF_05</i>	CpG_5	
<i>INSIGF_06</i>	CpG_6	
<i>KCNQ1OT1_01</i>	CpG_1	Mass overlap with unit 8
<i>KCNQ1OT1_02</i>	CpG_2	
<i>KCNQ1OT1_03</i>	CpG_3-5	
<i>KCNQ1OT1_04</i>	CpG_6	Mass overlap with units 15 and 19
<i>KCNQ1OT1_05</i>	CpG_7	
<i>KCNQ1OT1_06</i>	CpG_8&9	
<i>KCNQ1OT1_07</i>	CpG_10-12	Mass overlap with unit 3
<i>KCNQ1OT1_08</i>	CpG_13&14	
<i>KCNQ1OT1_09</i>	CpG_15	
<i>KCNQ1OT1_10</i>	CpG_16	Mass overlap with units 5 and 19
<i>KCNQ1OT1_11</i>	CpG_17&18	
<i>KCNQ1OT1_12</i>	CpG_19	
<i>KCNQ1OT1_13</i>	CpG_20	rs7940500
<i>KCNQ1OT1_14</i>	CpG_21	
<i>KCNQ1OT1_15</i>	CpG_22	
<i>KCNQ1OT1_16</i>	CpG_23	rs379976
<i>KCNQ1OT1_17</i>	CpG_24	
<i>KCNQ1OT1_18</i>	CpG_25	
<i>KCNQ1OT1_19</i>	CpG_26&27	Mass overlap with units 5 and 15
<i>MEG3_01</i>	CpG_1	Mass overlap with unit 6
<i>MEG3_02</i>	CpG_2	
<i>MEG3_03</i>	CpG_3	
<i>MEG3_04</i>	CpG_4	
<i>MEG3_05</i>	CpG_5	

Locus_CpG-unit^a	CpG-sites^a	Reason for removal prior to quality control
<i>MEG3_06</i>	CpG_6	Mass overlap with unit 5
<i>MEG3_07</i>	CpG_7	
<i>MEG3_08</i>	CpG_8&9	
<i>MEG3_09</i>	CpG_10&11	
<i>MEG3_10</i>	CpG_12-14	High Mass
<i>FTO_01</i>	CpG_1	Mass overlap with Unit 5
<i>FTO_02</i>	CpG_2&3	
<i>FTO_03</i>	CpG_4	Low mass
<i>FTO_04</i>	CpG_5	
<i>FTO_05</i>	CpG_6	Mass overlap with Unit 1
<i>FTO_06</i>	CpG_7	
<i>FTO_07</i>	CpG_8&9	
<i>FTO_08</i>	CpG_10&11	
<i>FTO_09</i>	CpG_12	Mass overlap with Unit 17
<i>FTO_10</i>	CpG_13	
<i>FTO_11</i>	CpG_14	
<i>FTO_12</i>	CpG_15	
<i>FTO_13</i>	CpG_16	
<i>FTO_14</i>	CpG_17	
<i>FTO_15</i>	CpG_18	
<i>FTO_16</i>	CpG_19	
<i>FTO_17</i>	CpG_20	Mass overlap with Unit 9
<i>APOC1_01</i>	CpG_1	
<i>APOC1_02</i>	CpG_2	
<i>APOC1_03</i>	CpG_3	
<i>APOC1_04</i>	CpG_4	
<i>APOC1_05</i>	CpG_5&6	rs402204
<i>APOC1_06</i>	CpG_7-9	High Mass and rs5111
<i>APOC1_07</i>	CpG_10	
<i>APOC1_08</i>	CpG_11	
<i>GNASAS_01</i>	CpG_1&2	
<i>GNASAS_02</i>	CpG_3&4	
<i>GNASAS_03</i>	CpG_5	
<i>GNASAS_04</i>	CpG_6	
<i>GNASAS_05</i>	CpG_7	

Locus_CpG-unit^a	CpG-sites^a	Reason for removal prior to quality control
<i>GNASAS_06</i>	CpG_8&9	
<i>GNASAS_07</i>	CpG_10-12	
<i>GNASAS_08</i>	CpG_13&14	
<i>GNASAS_09</i>	CpG_15	
<i>GNASAS_10</i>	CpG_16	rs45596642
<i>GNASAS_11</i>	CpG_17-19	
<i>GNAS A/B_01</i>	CpG_1	
<i>GNAS A/B_02</i>	CpG_2	Mass overlap with unit 4
<i>GNAS A/B_03</i>	CpG_3&4	
<i>GNAS A/B_04</i>	CpG_5&6	Mass overlap with unit 2
<i>GNAS A/B_05</i>	CpG_7	
<i>GNAS A/B_06</i>	CpG_8	
<i>GNAS A/B_07</i>	CpG_9&10	
<i>GNAS A/B_08</i>	CpG_11	Low mass
<i>GNAS A/B_09</i>	CpG_12	
<i>GNAS A/B_10</i>	CpG_13-15	
<i>GNAS A/B_11</i>	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

