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

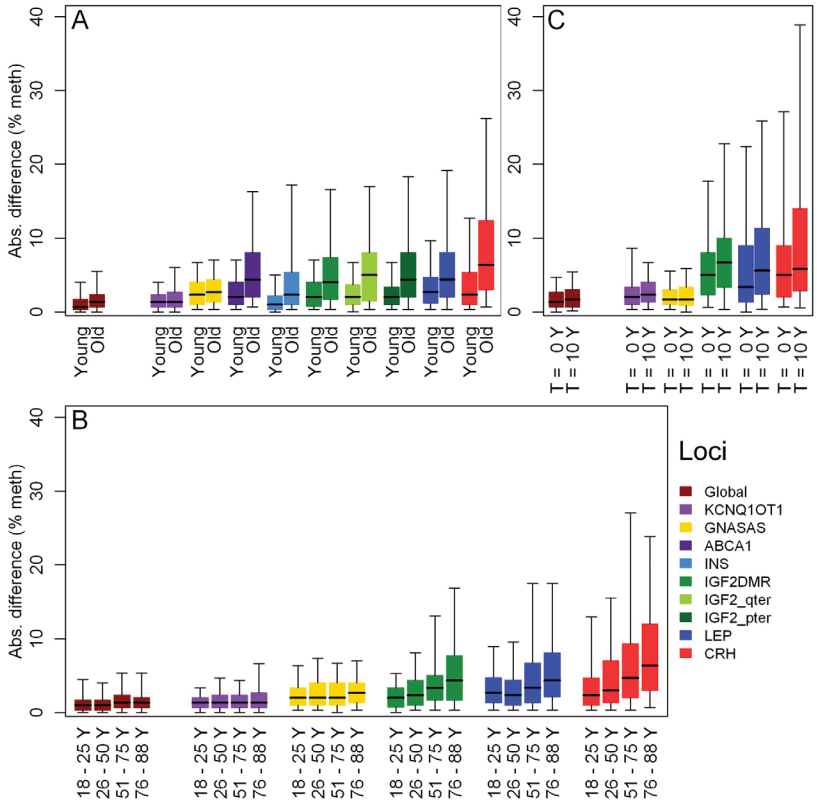


Figure 4.1: Increase in within-pair methylation discordance with age. Absolute within pair difference in % DNA methylation for global DNA methylation and for specific loci plotted for **A.** Young MZ twins (under 30 years old; n = 66 pairs) vs. old MZ twins (over 74 years old; n = 67 pairs). **B.** The 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.

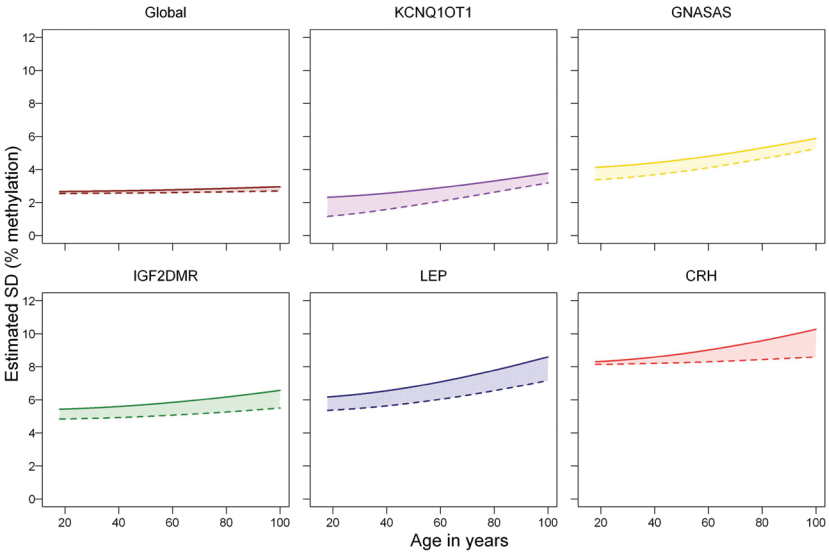


Figure 4.2: Contribution of familial and unique (individual) environmental factors to increasing methylation variation with age.
Estimated changes in the variation in DNA methylation (y-axis, given as standard deviation in percentage DNA methylation) plotted against the adult lifespan (x-axis, age in 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

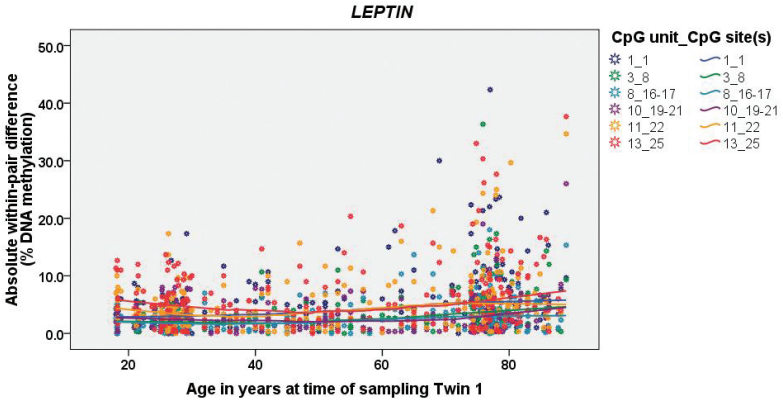
The linear mixed model uses all available methylation data per locus (assay), i.e. methylation of multiple CpG units per locus, accounts for the correlation between methylation of CpG units within a locus, and using this correlation handles data missing at random. It further enables the inclusion of 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.

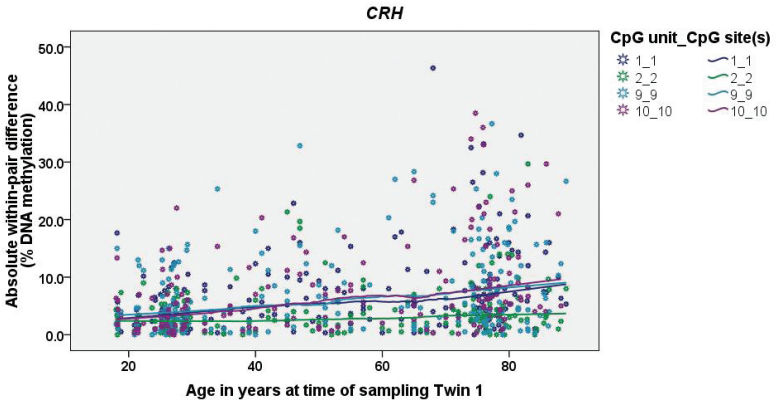
Variance component analysis in twin studies

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.

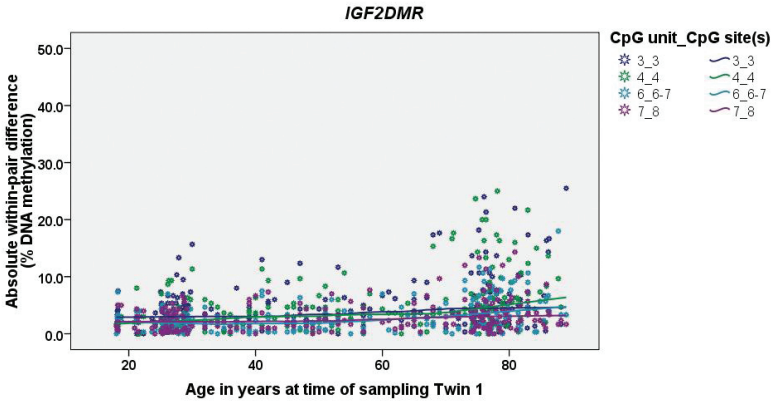
Supplementary Figures



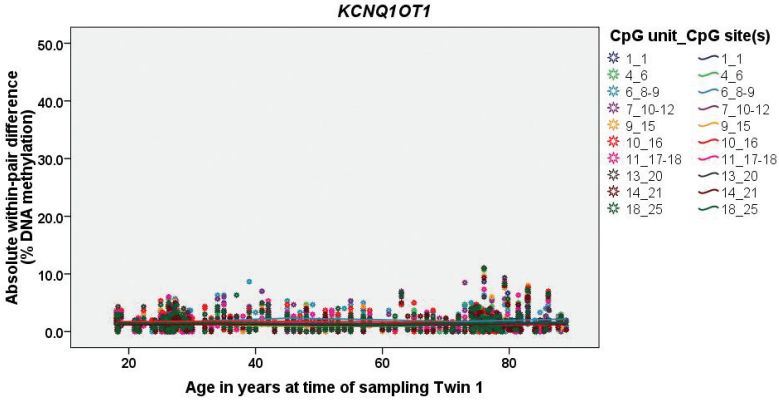
Supplementary Figure S4.1: The Absolute within MZ twin pair difference in % DNA methylation (y-axis) plotted against age (x-axis) per CpG unit of the *LEPTIN* locus. Each dot represents the absolute within pair difference of one MZ pair at one CpG unit. The different CpG units are represented by different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



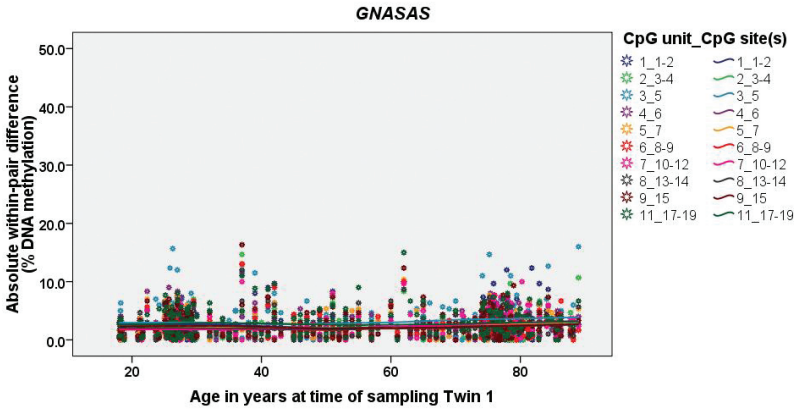
Supplementary Figure S4.2: The Absolute within MZ twin pair difference in % DNA methylation (y-axis) plotted against age (x-axis) per CpG unit of the *CRH* locus. Each dot represents the absolute within pair difference of one MZ pair at one CpG unit. The different CpG units are represented by different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



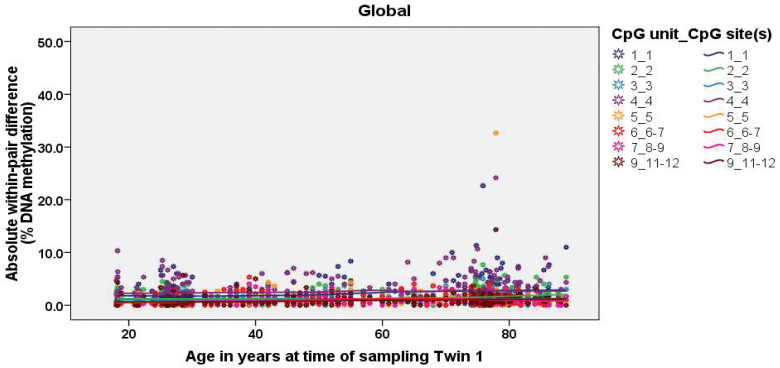
Supplementary Figure S4.3: The Absolute within MZ twin pair difference in % DNA methylation (y-axis) plotted against age (x-axis) per CpG unit of the *IGF2DMR* locus. Each dot represents the absolute within pair difference of one MZ pair at one CpG unit. The different CpG units are represented by different 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