Longitudinal change in function and biomarkers in human ageing

Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort

Qihua Tan\textsuperscript{1,2,*}, Bastiaan T Heijmans\textsuperscript{3}, Jacob v B Hjelmborg\textsuperscript{1}, Mette Soerensen\textsuperscript{1,2}, Kaare Christensen\textsuperscript{1,2} and Lene Christiansen\textsuperscript{1}

\textsuperscript{1}Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark, Odense C, Denmark, \textsuperscript{2}Unit of Human Genetics, Department of Clinical Research, University of Southern Denmark and Odense University Hospital, Odense C, Denmark and \textsuperscript{3}Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

\*Corresponding author. Epidemiology, Biostatistics and Biodemography, Department of Public Health, Faculty of Health Science, University of Southern Denmark, J. B. Winsløws Vej 9B, DK-5000, Odense C, Denmark. E-mail: qtan@health.sdu.dk

Accepted 29 April 2016

Abstract

\textbf{Background}: Current epigenetic studies on aging are dominated by the cross-sectional design that correlates subjects’ ages or age groups with their measured epigenetic profiles. Such studies have been more aimed at age prediction or building up the epigenetic clock of age rather than focusing on the dynamic patterns in epigenetic changes during the aging process.

\textbf{Methods}: We performed an epigenome-wide association study of intra-individual longitudinal changes in DNA methylation at CpG (cytosine-phosphate-guanine) sites measured in whole-blood samples of a cohort of 43 elderly twin pairs followed for 10 years (age at intake 73–82 years). Biological pathway analysis and survival analysis were also conducted on CpGs showing longitudinal change in their DNA-methylation levels. Classical twin models were fitted to each CpG site to estimate the genetic and environmental effects on DNA-methylation.

\textbf{Results}: Our analysis identified 2284 CpG sites whose DNA-methylation levels changed longitudinally over the follow-up. Twin modelling revealed that the longitudinal change for 90\% of these CpG sites was explained solely by individual unique environmental factors and only for 10\% of these sites was it influenced by familial factors (genetic or shared environment). Over 60\% of the identified CpG sites were replicated (same direction and replication $P < 0.05$) in an independent cross-sectional sample of 300 twins aged from 30 to 74 years. The replication rate went up to 91\% for the top 53 CpGs with $P < 1 \times 10^{-07}$. Pathway analysis of genes linked to these CpGs identified biologically meaningful gene-sets involved in cellular-signalling events and in transmission across chemical synapses, which are important molecular underpinnings of aging-related degenerative disorders.
Conclusion: Our epigenome-wide association studies on a cohort of old twins followed up for 10 years identified highly replicable epigenetic biomarkers predominantly implicated in signalling pathways of degenerative disorders and survival in the elderly.

Key words: DNA methylation, longitudinal, twins, aging, survival

Introduction
The aging process involves continuous remodelling of biological parameters from molecular modification to systems functional regulation through multiple biological mechanisms including epigenetics. The role of epigenetics in aging and age-related diseases is a key issue because certain epigenetic factors are thought to mediate, at least in part, the relationship between the environment and the genome. In fact, epigenetic modification in humans has been associated with age-related diseases including type 2 diabetes, cancer, cardiovascular diseases and neurodegenerative diseases. Recently, studying the direct relationship between epigenetic mechanisms and the aging process itself is gaining increasing attention. The potential reversibility of these epigenetic changes that occur as a hallmark of aging offers exciting opportunities to alter the trajectory of age-related diseases. This is especially important given the remarkable plasticity of aging. In the literature, age-associated epigenetic alterations have been identified by epigenome-wide association studies (EWAS) conducted on samples from different human tissues with blood, the most widely available and easily assessable DNA source, as a promising target tissue for studying the epigenetics of human aging.

Current epigenetic studies on aging are dominated by the cross-sectional design that correlates subjects’ ages or age groups with their measured epigenetic profiles. Such studies have been more aimed at age prediction or building up the epigenetic clock of age than focusing on the epigenetics of aging. In order to explore the trajectory of epigenetic regulation during the aging process, longitudinal observations in the aging subjects especially in the elderly are more preferable. In the literature, there have been a limited number of epigenetic association studies using longitudinal design. These studies were limited to a few candidate sites to very young individuals, e.g. from birth to 18 months and from 5 to 10 years or on samples with considerable age variation. Very recently, Shah et al. reported a longitudinal epigenetic study on elderly cohorts with a median age over 70 years and followed up for over 6 years. Their study revealed interesting genetic and environmental constraints on the longitudinal pattern of DNA methylation. Although interesting, their analysis focused on the stability of DNA methylation instead of intra-individual epigenetic drift over time.

We report results from a longitudinal EWAS performed on a cohort of 86 elderly twins (18 monozygotic or MZ pairs, 25 dizygotic or DZ pairs) with a mean age at intake of 76 years (range 73–82) who were followed up for a period of 10 years (from year 1997 to year 2007). By modelling the intra-individual longitudinal change in DNA methylation using the mixed-effect kinship model, we explored genomic sites that were differentially methylated over time and subsequently conducted a replication study in a large cross-sectional sample of 150 twin pairs aged 30–74 years. Survival analysis was carried out on CpGs showing longitudinal changes to associate their methylation levels at baseline and change in methylation over time with mortality. Taking advantage of our twin samples, we also report heritability estimates at CpG sites exhibiting intra-individual epigenetic drift over time.

Materials and methods
The study samples
The study samples consisted of 86 elderly Danish twins (18 monozygotic or MZ pairs, 25 dizygotic or DZ pairs)
collected by the Longitudinal Study of Aging Danish Twins initiated in 1995. The project collected like-sex twin pairs born in Denmark for longitudinal assessment of aging-related phenotypes. Our samples were collected from birth cohorts added in 1995 and in 1997 who were born before 1923 with age ranging from 73–82 years in 1997 when first blood samples were taken. The second blood samples were drawn in 2007 after a 10-year follow-up (Table 1). Individual survival information including age at death and censoring status (updated as of 2014) was collected from the Danish Central Population Register for survival analysis (bottom of Table 1). Zygosity information based on physical resemblance (extracted from the survey questioners) was further confirmed using 12 highly polymorphic microsatellite markers. Informed consent to participate in the study was obtained from all participants.

The samples for the replication study were cross-sectional samples of 150 pairs of identical or MZ twins aged 30–74 years (Supplementary Figure 3, available as Supplementary data at IJE online) with genome-wide DNA methylation measured using the same platform as in current longitudinal study (i.e. Illumina Infinium Human Methylation 450K BeadChip). As shown in Supplementary Figure 3 (available as Supplementary data at IJE online), the replication samples consist of two age groups, i.e. a younger group aged from 30–37 years (154 samples) and an older group aged 57–74 years (146 samples) with a gap of 20 years between the two groups. Both raw and processed DNA-methylation data for the replication samples have been deposited to the NCBI GEO database http://www.ncbi.nlm.nih.gov/geo/ under accession number GSE61496.

Both studies were approved by the Danish Scientific Ethics Committees and conducted in accordance with the Helsinki II declaration.

Blood sample collection and genomic DNA extraction

EDTA-anti-coagulated blood samples were collected and processed immediately or kept at room temperature for up to 24 hours, until further processing. The blood was centrifuged at 1000 g for 10 min, and buffy-coat was frozen in aliquots at −80°C. DNA was isolated from the buffy-coats using the salt precipitation method applying either a manual protocol or a semi-automated protocol based on the Autopure System (Qiagen, Hilden, Germany). Bisulphite treatment of 500 ng genomic DNA was carried out with the EZ-96 DNA-methylation kit (Zymo Research, Orange County, USA) following the manufacturer’s protocol.

Genome-wide DNA-methylation analysis

Genome-wide DNA methylation was measured using the Infinium Human Methylation 450K BeadChip (Illumina, San Diego, CA, USA) at the Leiden University Medical Center. The array allows simultaneous measurement of DNA-methylation status at 485 577 CpG sites across and beyond genes and CpG island regions in the human genome. The laboratory work for the arrays was performed according to the manufacturer’s instructions. Twins of each pair were processed together on the same array (each array contained 12 samples, i.e. 6 pairs of twins). Data normalization was done using the R package minfi which employs subset quantile within-array normalization for data pre-processing. At each CpG site, DNA-methylation level was summarized by calculating a ‘beta’ value defined by the Illumina’s formula as $\beta = M/(M + U + 100)$, where $M$ and $U$ are methylated ($M$) and unmethylated ($U$) signal intensities measured at the CpG site. The raw and processed DNA-methylation data have been deposited to the NCBI GEO database http://www.ncbi.nlm.nih.gov/geo/ under accession number GSE73115.

Quality control (QC)

Probe quality was controlled using the detection $P$-value defined as the proportion of samples reporting background signal levels for both methylated and unmethylated channels. The detection $P$-value was calculated by $minfi$. A $\beta$-value with its assigned detection $P$-value $> 0.01$ was treated as missing. CpG sites with more than 5% missing data were dropped from the analysis. Of the 485 577 CpGs measured by the array, 484 468 remained for subsequent analysis.
analysis after QC at probe level with 1109 CpGs removed. In addition, we also removed CpG sites harbouring SNPs (Single Nucleotide Polymorphisms), leaving 424 706 CpGs for subsequent analysis.

Estimating and adjusting cell composition
Cell composition in the whole blood can change as a result of aging. To deal with this problem, we estimated cell composition in each individual for six blood-cell types: CD8T, CD4T, natural killer cell, B cell, monocyte and granulocyte using the R package minfi. The package estimated cell composition based on the DNA-methylation data measured on a whole-blood sample and published cell-type-specific DNA-methylation data using an approach proposed by Houseman et al. Based on the estimates, the effect of cell composition was adjusted by including the change in cell composition over time as covariates in the regression model for statistical analysis.

Statistical analyses
Statistical model for longitudinal data
The longitudinal design was aimed at identifying and testing the longitudinal change in DNA methylation over the follow-up time. We used a mixed-effect kinship model modified from Tan et al. to regress longitudinal change on fixed (age at intake, sex, etc.) and random (MZ and DZ twin correlation) variables, i.e.

\[ \Delta \text{Meth} = \alpha + \beta_1 \text{age}_{97} + \beta_2 \text{sex} + \ldots + \text{Random effects.} \]  

(1)

Here, \( \Delta \text{Meth} = \logit(\text{Meth}_{07}) - \logit(\text{Meth}_{97}) \), with \( \text{Meth}_{07} \) and \( \text{Meth}_{97} \) as DNA-methylation levels measured in blood samples taken in 1997 and 2007. The fixed effect variable \( \text{age}_{97} \) is the intake age in 1997 when first blood sample was taken. In the model fitting, \( \text{age}_{97} \) was centred by subtracting the mean age. In this model, the intercept \( \alpha \) stands for the longitudinal change with an \( \beta_1 \) by subtracting the mean age. In this model, the intercept \( \alpha \) stands for the longitudinal change with an \( \beta_1 \) by subtracting the mean age. In this model, the intercept \( \alpha \) stands for the longitudinal change with an \( \beta_1 \) by subtracting the mean age.

Statistical model for cross-sectional data
Differently from the longitudinal data for discovery analysis, the data for the replication analysis were cross-sectional as described in the material section. Correspondingly, we applied a mixed-effect kinship model that regressed the mean DNA methylation on the fixed effects including age at blood collection, sex and age–sex interaction with random effect specified as twin pairing, i.e. we fitted

\[ \text{Meth} = \alpha + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{age*sex} + \ldots + \text{Random effects.} \]  

(2)

In this model, the effect of age measured by \( \beta_1 \) is equivalent to the longitudinal change measured by \( \alpha \) in the longitudinal model; the effect of age–sex interaction measured by \( \beta_3 \) is equivalent to the sex-specific longitudinal change in the longitudinal model. The effect of sex represents sex difference in the mean level of DNA methylation. Comparing the longitudinal with the cross-sectional models, one can easily see that the former avoids estimating the main effect of sex—an advantage that helps to improve the power for the longitudinal design. Similarly to the longitudinal model, a kinship matrix was also integrated into the covariance matrix of the cross-sectional model to fully adjust for the genetic correlation in the 150 MZ twin pairs. In model fitting, the methylation levels were logit transformed to ensure normal or approximately normal distribution.

Pathway analysis
The identified CpGs were annotated to nearest genes and evaluated for enrichment of gene-sets in the Reactome and the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways using Gene-Set Enrichment Analysis (GSEA) (http://www.broadinstitute.org/gsea/index.jsp).

Survival analysis
Based on individual survival information (age at death, censoring status), we fitted survival models to CpG sites displaying longitudinal change, in order to associate their methylation at baseline (year 1997) and change over time (1997–2007) with mortality. Considering the correlated structure of twin data, we introduced a Cox proportional hazard model with random effect (Gamma distributed frailty). The model included baseline and
intra-individual longitudinal change for DNA methylation as covariates adjusting for age at intake and sex, i.e. we fitted hazard for twin $j$ (1 or 2) in pair $i$ as

$$h_{ij}(t) = h_0(t)e^{\left(\beta_0 + \beta_1 \text{Meth}_{ij} + \beta_2 \text{Age}_{ij} + \beta_3 \text{Sex}_{ij} + v_i \right)}.$$  

(3)

In this model, $v_i$ is the unobserved random effect or frailty for the $ith$ pair following a Gamma distribution. The model estimated the effects on survival from baseline as well as from change in DNA methylation simultaneously.

All statistical analyses were performed using the R statistical software (http://www.r-project.org). The mixed-effect kinship model was fitted by applying the R package kinship (http://cran.r-project.org/src/contrib/Archive/kinship/). Correction for multiple testing in EWAS was done by estimating the false discovery rate (FDR) using the $p$.adjust() function in R. Survival models were fitted using the R package survival (http://cran.r-project.org/web/packages/survival/).

**Twin modelling**

The twin modelling was done by fitting the structural equation models to intra-individual longitudinal change in DNA-methylation level, i.e. $\Delta$Meth. For each phenotype, the variance and covariance of $\Delta$Meth were decomposed into the additive genetic (A), dominant genetic (D), common or shared environmental (C), and unique environmental (E) components in fitting the univariate twin models. Since the effects of C and D can be confounded in the classical twin study of MZ and DZ pairs reared together and cannot be estimated simultaneously, two separate models containing the A, C and E components (the ACE model) and the A, D and E components (the ADE model) were fitted. For each CpG site, the model fitting started with fitting both full ACE and full ADE models and the model with better performance indicated by lower Akaike Information Criterion (AIC). was chosen as the preferred full model for subsequent analysis. Based on the full ACE model, nested models were fitted by dropping the C (AE model), the A (CE model) or both (E model) components for best model selection. Likewise, three nested models (AE, DE and E) were fitted for comparison with the full ADE model. The likelihood ratio test (LRT) was applied for comparing the performances between the full model and its nested models. LRT calculates twice the difference in the log likelihoods between the full and the nested models which approximately follows the chi-squared distribution with the degree of freedom equals the difference in the number of parameters in the two models. In the model comparison, the parsimonious model was preferred when no statistical difference was observed between two models. The effects of age at intake and sex were adjusted by including them as covariates in the model fitting. Twin modelling was done by the R package mets (http://cran.r-project.org/web/packages/mets/index.html).

**Results**

**Discovery stage**

Our analysis identified 2284 CpGs out of the 424 706 analysed CpGs displaying intra-individual longitudinal change with a false discovery rate (FDR) $< 0.05$ and corresponding $P$-value $< 2.75 \times 10^{-04}$ (Supplementary Table 1 and Supplementary Figure 1, available as Supplementary data at IJE online). In Table 2, we show the top 53 CpGs with $P < 1 \times 10^{-07}$. Three CpGs listed in Table 2 (cg24724428: $P = 8.81 \times 10^{-12}$, cg16867657: $P = 6.35 \times 10^{-10}$ and cg21572722: $P = 4.25 \times 10^{-08}$) were located within 1500 base pairs of the transcription start site of the ELOVL2 gene on chromosome 6. There were no CpGs showing sex-specific longitudinal changes with FDR $< 0.05$ (Supplementary Table 1, available as Supplementary data at IJE online).

Figure 1 is a scatter plot showing intra-individual longitudinal change for the top 10 CpGs. The age trajectory goes up or down at different levels with no obvious difference between the two sexes. Figure 2 (left-hand panel) is a volcano plot showing the $P$-values (in log scale) for all 424 706 CpGs plotted against corresponding regression coefficients. The CpGs associated with intra-individual longitudinal change with FDR $< 0.05$ are coloured in red (positive coefficients meaning increased methylation level during the follow-up) and green (negative coefficients meaning decreased methylation level during the follow-up). The symmetric pattern indicates that there was no predominant trend of up or down regulation on DNA methylation during aging. In the right-hand panel, we plot the $P$-values (in log scale) against the regression coefficients for sex (i.e. the sex-specific effect in the longitudinal model, see method section). The CpGs in the left-hand panel were plotted again with red colour in the right-hand panel using large empty dots. Differently from the left-hand panel, no CpG displayed a sex-specific effect with FDR $< 0.05$.

**Pathway analysis**

The 2284 CpGs in Supplementary Table 1 (available as Supplementary data at IJE online) were annotated to the
<table>
<thead>
<tr>
<th>CpG sites</th>
<th>Annotation</th>
<th>MAPINFO</th>
<th>UCSC_RefGene_Name</th>
<th>Relation_to_UCSC_CpG_Island</th>
<th>UCSC_RefGene_Coef.</th>
<th>UCSC_RefGene_P-value</th>
<th>Discovery Coef.</th>
<th>Discovery P-value</th>
<th>Replication Coef.</th>
<th>Replication P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg13823169</td>
<td>9</td>
<td>139776893</td>
<td>N_Shelf</td>
<td>Island</td>
<td>0.113</td>
<td>2.85E-12</td>
<td>-0.007</td>
<td>9.03E-17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg24724428</td>
<td>6</td>
<td>11044888</td>
<td>ELOVL2 Island</td>
<td>TSS1500</td>
<td>0.209</td>
<td>8.81E-12</td>
<td>0.017</td>
<td>2.51E-34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg04875128</td>
<td>15</td>
<td>31775895</td>
<td>OTUD7A Island</td>
<td>Island</td>
<td>0.248</td>
<td>5.39E-11</td>
<td>0.036</td>
<td>2.27E-35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg04208403</td>
<td>16</td>
<td>49525807</td>
<td>ZNF423 N_Shore</td>
<td>Body</td>
<td>-0.094</td>
<td>1.20E-10</td>
<td>-0.006</td>
<td>1.59E-24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg06784991</td>
<td>1</td>
<td>53308768</td>
<td>ZYG11A Island</td>
<td>Body</td>
<td>0.127</td>
<td>3.73E-10</td>
<td>0.010</td>
<td>8.84E-26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg06679990</td>
<td>12</td>
<td>57863266</td>
<td>GLI1</td>
<td>Body</td>
<td>0.124</td>
<td>4.03E-10</td>
<td>-0.001</td>
<td>0.625383</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg16867657</td>
<td>6</td>
<td>11044877</td>
<td>ELOVL2 Island</td>
<td>Island</td>
<td>0.179</td>
<td>6.35E-10</td>
<td>0.023</td>
<td>3.28E-86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg17269633</td>
<td>17</td>
<td>42736300</td>
<td>C17orf104 TSS1500</td>
<td>Island</td>
<td>0.209</td>
<td>8.81E-12</td>
<td>0.009</td>
<td>2.77E-09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg23718736</td>
<td>18</td>
<td>6413908</td>
<td>L3MBTL4 Island</td>
<td>TSS1500</td>
<td>0.017</td>
<td>9.03E-17</td>
<td>0.005</td>
<td>9.05E-05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg19856758</td>
<td>2</td>
<td>62684655</td>
<td>S_Shore</td>
<td>Body</td>
<td>-0.014</td>
<td>1.78E-09</td>
<td>0.017</td>
<td>4.18E-22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg14106195</td>
<td>5</td>
<td>76464212</td>
<td>N_Shore</td>
<td>Island</td>
<td>-0.125</td>
<td>1.97E-09</td>
<td>-0.017</td>
<td>0.49E-22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg18997983</td>
<td>5</td>
<td>1109031</td>
<td>S_Shore</td>
<td>Body</td>
<td>-0.129</td>
<td>2.49E-09</td>
<td>0.003</td>
<td>0.969662</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg00081372</td>
<td>5</td>
<td>15398923</td>
<td>S_Shore</td>
<td>Body</td>
<td>-0.138</td>
<td>2.68E-09</td>
<td>-0.003</td>
<td>0.03185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg26186732</td>
<td>22</td>
<td>16165225</td>
<td>Island</td>
<td>TSS1500</td>
<td>0.11</td>
<td>3.13E-09</td>
<td>0.005</td>
<td>0.004203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg06400319</td>
<td>1</td>
<td>153045274</td>
<td>SPRR2B</td>
<td>Island</td>
<td>0.179</td>
<td>6.35E-10</td>
<td>0.023</td>
<td>3.28E-86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg17269633</td>
<td>17</td>
<td>42736300</td>
<td>C17orf104 TSS1500</td>
<td>Island</td>
<td>0.209</td>
<td>8.81E-12</td>
<td>0.009</td>
<td>2.77E-09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg23718736</td>
<td>18</td>
<td>6413908</td>
<td>L3MBTL4 Island</td>
<td>TSS1500</td>
<td>0.017</td>
<td>9.03E-17</td>
<td>0.005</td>
<td>9.05E-05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg04875128</td>
<td>15</td>
<td>31775895</td>
<td>OTUD7A Island</td>
<td>Body</td>
<td>-0.094</td>
<td>1.20E-10</td>
<td>-0.006</td>
<td>1.59E-24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg06784991</td>
<td>1</td>
<td>53308768</td>
<td>ZYG11A Island</td>
<td>Body</td>
<td>0.127</td>
<td>3.73E-10</td>
<td>0.010</td>
<td>8.84E-26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg06679990</td>
<td>12</td>
<td>57863266</td>
<td>GLI1</td>
<td>Body</td>
<td>0.124</td>
<td>4.03E-10</td>
<td>-0.001</td>
<td>0.625383</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1358 nearest genes and evaluated for enrichment of gene-sets in Reactome and the KEGG pathways using GSEA. In the Reactome database, a total of 64 biological pathways were identified with FDR < 0.01 (Supplementary Table 2, available as Supplementary data at IJE online). In the KEGG database, 57 biological pathways had FDR < 0.01 (Supplementary Table 3, available as Supplementary data at IJE online). In Table 3, we show the top 10 pathways from each database. The top 10 Reactome pathways are involved in signalling transmission and cellular communication,
developmental biology and metabolism of lipids and lipoproteins. Nearly all of the top 10 KEGG pathways are involved in cellular-signalling activities and processes.

**Replication stage**

In order to reconfirm the findings from our longitudinal samples, we conducted a replication study on a relatively large cross-sectional dataset of 300 samples from 150 pairs of birth-weight-discordant monozygotic twins aged 30–74 years (Supplementary Figure 2, available as Supplementary data at IJE online) with genome-wide DNA methylation measured using the same platform as in current study (i.e. Illumina Infinium Human DNA methylation 450K bead-chip). The data were analysed by Tan et al.27 who reported no epigenetic association with birth-weight discordance.

The mixed-effect kinship model for cross-sectional data (see the ‘Methods’ section) was applied to associate DNA methylation with age while adjusting for sex, birth weight, blood-cell composition and accounting for twin-pair correlation.

Of the 2284 CpGs associated with intra-individual longitudinal methylation changes in the discovery sample, 1404 (61%) CpGs were replicated with consistently the same direction and with \( \hat{P} < 0.05 \) (Supplementary Table 1, right-hand side, available as Supplementary data at IJE online). For the 53 CpGs with \( \hat{P} < 1 \times 10^{-07} \) in Table 2, 48 were replicated with the same direction and \( \hat{P} < 0.05 \), and 33 CpGs even with comparable \( \hat{P} \)-values with \( \hat{P} < 2.68 \times 10^{-08} \), i.e. a replication rate as high as 91%, suggesting that the top CpGs are highly replicable.

**Table 3. The top 10 REACTOME and KEGG pathways over-represented by the significant CpGs**

<table>
<thead>
<tr>
<th>Gene-set name [# genes (K)] Description</th>
<th>Overlap k</th>
<th>( P )-value</th>
<th>FDR q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>REACTOME pathways</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEURONAL_SYSTEM [279] Genes involved in neuronal system</td>
<td>45</td>
<td>( 2.1 \times 10^{-20} )</td>
<td>( 1.42 \times 10^{-17} )</td>
</tr>
<tr>
<td>TRANSMISSION_ACROSS_CHEMICAL_SYNAPSES [186] Genes involved in transmission across chemical synapses</td>
<td>29</td>
<td>( 2.86 \times 10^{-13} )</td>
<td>( 9.63 \times 10^{-11} )</td>
</tr>
<tr>
<td>DEVELOPMENTAL_BIOLOGY [396] Genes involved in developmental biology</td>
<td>41</td>
<td>( 5.07 \times 10^{-12} )</td>
<td>( 8.78 \times 10^{-10} )</td>
</tr>
<tr>
<td>HEMOSTASIS [466] Genes involved in hemostasis</td>
<td>45</td>
<td>( 5.21 \times 10^{-12} )</td>
<td>( 8.78 \times 10^{-10} )</td>
</tr>
<tr>
<td>AXON_GUIDANCE [251] Genes involved in axon guidance</td>
<td>31</td>
<td>( 2.33 \times 10^{-11} )</td>
<td>( 3.15 \times 10^{-9} )</td>
</tr>
<tr>
<td>NEUROTNSFER_RECEPTOR_BINBINDING_AND_DOWNSTREAM_TRANSMISSIO-N_IN_THE_POSTSYNAPTIC_CELL [137] Genes involved in neurotransmitter receptor binding and downstream transmission in the post-synaptic cell</td>
<td>22</td>
<td>( 1.14 \times 10^{-10} )</td>
<td>( 1.28 \times 10^{-8} )</td>
</tr>
<tr>
<td>SIGNALING_BY_GPCR [920] Genes involved in signalling by GPCR</td>
<td>64</td>
<td>( 3.64 \times 10^{-10} )</td>
<td>( 3.51 \times 10^{-8} )</td>
</tr>
<tr>
<td>GPCR_LIGAND_BINDING [408] Genes involved in GPCR ligand binding</td>
<td>37</td>
<td>( 2.25 \times 10^{-9} )</td>
<td>( 1.89 \times 10^{-7} )</td>
</tr>
<tr>
<td>PLATELET_ACTIVATION_SIGNALING_AND_AGGREGATION [208] Genes involved in Platelet activation, signalling and aggregation</td>
<td>25</td>
<td>( 3.43 \times 10^{-9} )</td>
<td>( 2.57 \times 10^{-7} )</td>
</tr>
<tr>
<td>METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS [478] Genes involved in metabolism of lipids and lipoproteins</td>
<td>40</td>
<td>( 5.15 \times 10^{-9} )</td>
<td>( 3.47 \times 10^{-7} )</td>
</tr>
<tr>
<td>KEGG pathways</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION [272] Neuroactive ligand–receptor interaction</td>
<td>34</td>
<td>( 1.82 \times 10^{-12} )</td>
<td>( 3.38 \times 10^{-10} )</td>
</tr>
<tr>
<td>FOCAL_ADHESION [201] Focal adhesion</td>
<td>28</td>
<td>( 1.2 \times 10^{-11} )</td>
<td>( 1.11 \times 10^{-9} )</td>
</tr>
<tr>
<td>CALCIUM_SIGNALING_PATHWAY [178] Calcium-signalling pathway</td>
<td>25</td>
<td>( 1.25 \times 10^{-10} )</td>
<td>( 7.72 \times 10^{-9} )</td>
</tr>
<tr>
<td>FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS [97] Fc gamma R-mediated phagocytosis</td>
<td>18</td>
<td>( 4.88 \times 10^{-10} )</td>
<td>( 1.87 \times 10^{-8} )</td>
</tr>
<tr>
<td>MAPK_SIGNALING_PATHWAY [267] MAPK-signalling pathway</td>
<td>30</td>
<td>( 5.02 \times 10^{-10} )</td>
<td>( 1.87 \times 10^{-8} )</td>
</tr>
<tr>
<td>REGULATION_OF_ACTIN_CYTOSKELETON [216] Regulation of actin cytoskeleton</td>
<td>26</td>
<td>( 1.62 \times 10^{-9} )</td>
<td>( 5.03 \times 10^{-8} )</td>
</tr>
<tr>
<td>PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM [76] Phosphatidylinositol signalling system</td>
<td>15</td>
<td>( 5.56 \times 10^{-9} )</td>
<td>( 1.48 \times 10^{-7} )</td>
</tr>
<tr>
<td>LONG_TERM_POTENTIATION [70] Long-term potentiation</td>
<td>14</td>
<td>( 1.49 \times 10^{-8} )</td>
<td>( 3.47 \times 10^{-7} )</td>
</tr>
<tr>
<td>CHEMOKINE_SIGNALING_PATHWAY [190] Chemokine signalling pathway</td>
<td>22</td>
<td>( 5.77 \times 10^{-8} )</td>
<td>( 1.19 \times 10^{-6} )</td>
</tr>
<tr>
<td>FC_EPSILON_RI_SIGNALING_PATHWAY [79] Fc epsilon RI signalling</td>
<td>14</td>
<td>( 7.51 \times 10^{-8} )</td>
<td>( 1.4 \times 10^{-6} )</td>
</tr>
</tbody>
</table>
Supplementary Table 1 (available as Supplementary data at IJE online). In contrast, for the effect on survival from the baseline DNA-methylation level in 1997, 78 had $P < 0.01$, 5 had $P < 1 \times 10^{-03}$ and none had $P < 1 \times 10^{-04}$. Finally, the associations with survival from the longitudinal change and from the baseline level of DNA methylation are exclusive to each other meaning that the same CpG site did not exhibit survival association both with the longitudinal change and with the baseline DNA-methylation level. Only two CpGs reached FDR < 0.05 for effect on survival by longitudinal methylation change, no CpG showed similar FDR level for survival by baseline methylation.

**Heritability estimates**

In Supplementary Table 4 (available as Supplementary data at IJE online), we show the parameter estimates from the twin models fitted to each CpG site for the proportions of total variances in intra-individual longitudinal change (left-hand side), in the baseline methylation (middle) and at the end of follow-up (right-hand side), explained by the additive genetic (A), dominant genetic (D), shared environmental (C) and unique environmental (E) factors together with their 95% CIs. For the longitudinal change at the individual methylation level, as many as 89.9% (2053 out of 2284) of the CpGs were best fitted by the model that contained only the E component, i.e. the E model, 5.5% (126 CpGs) fitted by the DE model, 3.2% (73 CpGs) by the CE model and only 1.4% (32 CpGs) by the AE model. In the models, the additive genetic component A accounted for, on average, 36% (range: 27.1–70.5%) of the total variance in the longitudinal change in the AE model, the shared family environment C for 26.9% (range: 21.4–38.2%) in the CE model, the dominant genetic effect D for 41.6% (range: 27.2–75.3%) in the DE model and the unique individual environment E for predominantly the major part of the total variation in all models. In contrast to the longitudinal change, the A, D and C components were largely increased in the total variances of DNA methylation at intake and at end of follow-up (Supplementary Table 4, middle and right-hand side, available as Supplementary data at IJE online). For the methylation level at intake, the percentages of ACE/AE, DE, CE and E models were 15.8%, 22%, 19.5% and 42.7% whereas, for the level at end of follow-up, they were 13.4%, 24.6%, 14.5% and 47.5%.

To determine the degree of additivity or dominance of the genetic effects, and the relative contributions of the environmental factors, we fitted the twin models to each CpG site. In Supplementary Table 4 (available as Supplementary data at IJE online), we illustrate the different patterns of estimates for the twin models fitted to longitudinal change (top), level at intake (middle) and at the end of trace (bottom), with the x-axis for CpG sites arranged according to their best-fitting models ACE/AE, DE, CE and E sequentially but fitted to the intake level. The figure reveals a...
remarkable change in the genetic and environmental contributions to the levels of DNA methylation after 10 years in the elderly cohort. Despite the different patterns, all three figures illustrate the high importance of a unique environment in regulating the level and especially the longitudinal change of DNA methylation in the elderly.

**Discussion**

We have conducted a longitudinal EWAS in aging twins of intra-individual longitudinal changes in DNA methylation characterized by the following features:

1. A targeted analysis of epigenetic drift at advanced ages in an elderly cohort followed up longitudinally for 10 years with baseline age from 73 to 82 years—an age period that has been rarely touched by previous studies.
2. A powerful longitudinal design focused on intra-individual epigenetic variation over time instead of looking at difference between samples of different ages, enabling the control of potential confounding factors that could affect DNA-methylation status.
3. A strong replication study using a large independent sample of complete different design, i.e. the cross-sectional design enabling stringent replication and validation of the discovery results.
4. A survival analysis on CpGs sites showing longitudinal change to assess their effects on individual survival from the baseline DNA-methylation level as well as from the altered DNA methylation over time.
5. A twin modelling on DNA-methylation phenotype to examine the genetic and environmental influence on longitudinal epigenetic change at individual CpG sites.

Advantaged by the above-mentioned features, our EWAS was able to identify CpG sites displaying longitudinal changes which were highly replicable (same direction, \(P < 0.05\)) with a replication rate of 61% for CpGs with FDR < 0.05 in the discovery phase. For the 53 top CpGs with \(P < 1 \times 10^{-07}\), the replication rates were as high as 91%, with the majority of them (33 CpGs) replicated at the same level of \(P\)-value as in the discovery stage, with \(P < 1 \times 10^{-07}\) (Table 2). Note that the high replication rate does not mean that the cross-sectional design can be equally efficient as the longitudinal design given the fact that the size of cross-sectional replication samples (300) is much larger than that of the longitudinal samples (86). Although making inference on the mean of individual changes (longitudinal design) or on the difference in means (cross-sectional design) should theoretically arrive at the same conclusion, the former is advantaged by the ability to control confounding factors which add noise to the latter. As such, the cross-sectional design is less powerful than the longitudinal design and requires large sample sizes to overcome the disadvantage. Overall, the high replication rate in a sample of completely different study design and different age ranges could suggest (i) high reliability of our identified CpGs and (ii) the aging-associated epigenetic modification could have already started from much earlier ages.

In a recent meta-analysis, Bacalini et al. compiled a list of 44 age-associated epigenomic regions and their linked genes. Among them, the ELOVL2 gene on chromosome 6 had 3 CpGs in our list of 53 CpGs (\(P < 1 \times 10^{-07}\)) in Table 2: cg24724428 (\(P = 8.81 \times 10^{-12}\)), cg16687657 (\(P = 6.35 \times 10^{-10}\)) and cg21572722 (\(P = 4.25 \times 10^{-08}\)); all were hypermethylated with aging. The ELOVL2 gene has been frequently reported to associate with age or aging from multiple studies conducted in different European populations including Italian, Swedish, German, Polish as well as in American Caucasian samples. It is interesting to see that the three CpGs were nicely replicated for their association with age in the cross-sectional samples as shown in the right-hand side of Table 2 (cg16687657, \(P = 3.28 \times 10^{-86}\); cg21572722, \(P = 1.29 \times 10^{-54}\); cg24724428, \(P = 2.51 \times 10^{-34}\)). Since both our discovery and replication samples were Danish twins, this study confirms the association of ELOVL2 gene with age (based on our cross-sectional samples) and aging (based on our longitudinal samples) in the Danish population both in the discovery and in the replication stages.

Another interesting CpG site from the candidate regions of Bacalini et al. was annotated to the OTUD7A gene on chromosome 15 with two CpGs in Table 1: cg04878128 (\(P = 5.39 \times 10^{-11}\)) and cg01763090 (\(P = 5.91 \times 10^{-08}\)); both were hypermethylated with aging reinforcing the result on the OTUD7A gene by Hannum et al. The association with aging at the two CpG sites was highly replicated by our cross-sectional samples, with \(P = 2.27 \times 10^{-35}\) for cg04875128 and \(P = 2.35 \times 10^{-15}\) for cg01763090. One more interesting gene from the list of Bacalini et al. is ZYG11A on chromosome 1, with one CpG listed at the top of Table 2, cg06784991 (\(P = 3.73 \times 10^{-10}\)), displaying increased methylation level with aging. The DNA-methylation status of the ZYG11A gene has been shown to correlate with age by Florath et al. The association with aging identified by our longitudinal samples was again replicated by our cross-sectional samples reinforcing the increasing methylation pattern with age (\(P = 8.84 \times 10^{-26}\)). Besides the ELOVL2, OTUD7A and ZYG11A genes, there were three more genes (EDARADD, GPR78 and PRLHR) from Table 1 (\(P < 4.62 \times 10^{-08}\)) that overlapped with the candidate regions from the list of Bacalini et al.; all three sites were replicated by our cross-sectional data,
with \( P < 6.94 \times 10^{-17} \). Finally, we examined the overlaps between the 44 candidate sites and the 1358 unique genes annotated to the list of CpGs in Supplementary Table 1 (available as Supplementary data at IJE online). We found 21 out of the 44 sites in the CpGs identified by our longitudinal samples—a coverage of 48%. In Figure 3, the large solid dots represent the 21 CpGs annotated to the candidate sites of Bacalini et al.\(^{37}\) Their associations with age are very well captured by our discovery and replication samples.

Among the top 10 most highly enriched canonical pathways identified from the Reactome and the KEGG pathway databases (Table 3), many belong to the cellular-signalling events (e.g. the mitogen-activated protein kinase or MAPK signalling pathway, calcium-signalling pathway, focal adhesion, regulation of actin cytoskeleton) that govern basic cellular activities and coordinate cell actions. The MAPK pathways are highly conserved signalling pathways that regulate diverse cellular functions including cell proliferation, differentiation, migration and apoptosis.\(^{41}\) Regulations of the MAPK family have been linked to the inhibitory actions of excessive oxidative insult on adrenal steroid hormone production caused by aging,\(^{42}\) to skeletal muscle aging\(^{43,44}\) and to human neutrophil apoptosis.\(^{45}\) In the literature, the pathologica roles of MAPK-signalling pathways in human diseases have been intensively studied with a focus on cancer and neurodegenerative diseases including Alzheimer’s disease and Parkinson’s disease.\(^{46}\) Our result further reflects the high importance of the MAPK pathways in the human aging process. Calcium is a ubiquitous intracellular messenger controlling a diverse range of cellular processes such as gene transcription, muscle contraction and cell proliferation.\(^{47}\) Dysregulated expression of calcium-signalling pathways has been shown to occur with progression of Alzheimer-type pathology in the aging brain\(^{48,49}\) and contributes to an increased susceptibility to the degenerative process.\(^{50}\) Focal adhesions encompass integrin adhesion, signalling and the actin cytoskeleton to form mechanical links between intracellular actin bundles and the extracellular substrate in many cell types. Altered focal adhesion signalling has been linked to neuronal dystrophy and synaptic loss in the course of Alzheimer’s disease.\(^{51}\)

Interestingly, a biological pathway analysis from our very recent genome-wide association study of longevity in Han Chinese (2178 centenarians and 2299 middle-aged controls) identified MAPK and calcium-signalling pathways as highly associated with human longevity (unpublished data). Moreover, the two pathways were also found differentially regulated during aging in Caenorhabditis elegans in a microarray gene expression study.\(^{52}\) The consistent biological pathways found from genetic, epigenetic and transcriptomic approaches emphasize the important role of cellular-signalling events in the aging process and in affecting human survival.

Multiple biological pathways in Table 3, e.g. neuroactive ligand-receptor interaction, genes involved in transmission across chemical synapses, long-term potentiation, etc. encompass the neurochemistry of synapses. Normal aging appears to result in significant but restricted neurochemical changes in synapses. There is accumulating evidence that age-related decline in synaptic function in the brains contribute to changes in motor function, mood and memory, etc. In the literature, the effects of aging on long-term potentiation are well documented. Aging-related impairment in the strength of synapses that connect different cells within a neuronal network can lead to neurodegenerative disorders such as Alzheimer’s disease that cause marked cognitive decline and dementia.\(^{53}\) Overall, considering the fact that blood is a promising surrogate for brain tissue,\(^{12}\) the results from the pathway analysis could suggest that our identified CpG sites might represent important molecular underpinnings of aging-related degenerative disorders.

Our survival analysis only focused on the 2284 CpGs manifesting longitudinal change in DNA methylation over the 10 years’ follow-up with the aim of checking whether longitudinal methylation patterns were more associated with mortality as compared with the baseline methylation levels. Although only two CpGs had FDR < 0.05, our exploratory results showed higher importance of intra-individual longitudinal change in affecting individual survival than the effect by baseline methylation level at intake. Here, we want to point out that, even for longitudinal change in DNA methylation, only a limited number of CpGs were associated with survival, perhaps due to the high proportion of censoring (28%) in a relatively small sample.

The results from our twin modelling indicated that the intra-individual longitudinal change in DNA methylation at most of the CpG sites (89.9%) was solely controlled by individual unique or stochastic environmental factors (Supplementary Table 4 and Supplementary Figure 3, available as Supplementary data at IJE online). Longitudinal variation in DNA methylation at a small proportion of the CpG sites was regulated by genetic mechanisms (additive or dominant) with moderate mean contributions and large ranges of heritability estimate from low to high. In the 73 CpGs fitted to the CE model for longitudinal change, the contribution by a shared environment was very low (mean 26.9%, range: 21.4–38.2%), suggesting that the familial influence on the longitudinal patterns of DNA methylation at old ages is very limited. Although the proportion of CpGs regulated by A, D and C
components increased considerably in the twin models fitted to levels of DNA methylation at intake and at end of follow-up (Supplementary Figure 3, available as Supplementary data at IJE online), the predominant role of unique environment E applies both to the levels of methylation and especially to the longitudinal change in DNA methylation. In conclusion, unique or stochastic environmental factors could play a major role in modifying the trajectory of DNA methylation in the elderly.

Our current study focused on EWAS of aging without consideration of individual environmental factors. However, the high importance of individual unique environments in determining the aging-dependent patterns of DNA methylation as revealed by our twin modelling calls for well-designed sizeable epigenetic studies to look for major environmental factors involved, including individual behaviour or lifestyle (e.g. diet, smoking, drinking, sporting), social environment (e.g. education, occupation, sedentary work) and physical environment (e.g. living condition, workplace, water and air pollution). Tan et al.\textsuperscript{54} proposed the use of twins in epigenetic studies of aging to control for individual differences in genetic make-ups while focusing on identification of epigenetic alterations mediated by specific environmental factors. Differently from traditional epidemiology that associates environmental factors indirectly with health outcomes, such an approach provides a direct link that associates environmental exposure with molecular-level epigenetic modification and with health (here the aging process). Our current study is further limited by the fact that DNA-methylation data were collected for only two time points in a relatively long follow-up. Longitudinal EWAS with epigenetic measurements collected over multiple time points should provide more informative data concerning the dynamic patterns of epigenetic regulation during the aging process.

**Funding**

This work was supported by the Integrated research on Developmental determinants of Aging and Longevity (IDEAL), an EU FP7 project with number: 259679. The Danish Twin Registry is supported by a grant from the National Program for Research Infrastructure 2007 from the Danish Agency for Science Technology and Innovation.

**Acknowledgements**

We thank the IT support at our department for their excellent work that helped with computation and data management.

**Conflict of interest:** None declared.

**References**

42. Abidi P, Leers-Sucheta S, Cortez Y, Han J, Azhar S. Evidence that age-related changes in p38 MAP kinase contribute to the decreased steroid production by the adrenocortical cells from old rats. *Aging Cell* 2008;7:168–78.