GeroScience (2018) 40:469–484
https://doi.org/10.1007/s11357-018-0040-0

ORIGINAL ARTICLE

DNA methylation associated with healthy aging of elderly twins

Sangkyu Kim · Jennifer Wyckoff · Anne-T Morris · Annemarie Succop · Ally Avery · Glen E. Duncan · S. Michal Jazwinski

Received: 25 May 2018 / Accepted: 9 August 2018 / Published online: 22 August 2018
© American Aging Association 2018

Abstract Variation in healthy aging and lifespan is ascribed more to various non-genetic factors than to inherited genetic determinants, and a major goal in aging research is to reveal the epigenetic basis of aging. One approach to this goal is to find genomic sites or regions where DNA methylation correlates with biological age. Using health data from 134 elderly twins, we calculated a frailty index as a quantitative indicator of biological age, and by applying the Infinium HumanMethylation450K BeadChip technology to their leukocyte DNA samples, we obtained quantitative DNA methylation data on genome-wide CpG sites. We analyzed the health and epigenome data by taking two independent associative approaches: the parametric regression-based approach and a non-parametric machine learning approach followed by GO ontology analysis. Our results indicate that DNA methylation at CpG sites in the promoter region of *PCDHGA3* is associated with biological age. *PCDHGA3* belongs to clustered protocadherin genes, which are all located in a single locus on chromosome 5 in human. Previous studies of the clustered protocadherin genes showed that (1) DNA methylation is associated with age or age-related phenotypes; (2) DNA methylation can modulate gene expression; (3) dysregulated gene expression is associated with various pathologies; and (4) DNA methylation patterns at this locus are associated with adverse lifetime experiences. All these observations suggest that DNA methylation at the clustered protocadherin genes, including *PCDHGA3*, is a key mediator of healthy aging.

Keywords DNA methylation · Frailty index · Biological age · Protocadherin · Aging

Introduction

Aging is accompanied by functional deterioration and increasing risk for pathology and mortality. Estimates of the heritability of human longevity range from 0.15 to

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11357-018-0040-0) contains supplementary material, which is available to authorized users.

S. Kim (✉) · J. Wyckoff · S. M. Jazwinski
Tulane Center for Aging and Department of Medicine, Tulane University Health Sciences Center, New Orleans, LA 70112, USA
e-mail: skim5@tulane.edu

A.-T. Morris
Virginia Commonwealth University, Mid-Atlantic Twin Registry, Richmond, VA, USA

A. Succop · A. Avery · G. E. Duncan
University of Washington Twin Registry, Seattle, WA, USA

Present Address:
A. Avery · G. E. Duncan
Washington State Twin Registry, Washington State University – Health Sciences Spokane, Spokane, WA, USA

0.35 (Herskind et al. 1996; Gudmundsson et al. 2000; Kerber et al. 2001; Mitchell et al. 2001). This indicates that human aging varies much more with non-genetic factors. Indeed, monozygotic (MZ) twin pairs often show high discordance in many phenotypes, including age-related diseases (Castillo-Fernandez et al. 2014; Willemsen et al. 2015).

Epigenetic modifiers are important modulators of gene expression without changing DNA sequence. DNA methylation is important for differential gene expression during embryonic development (Meissner et al. 2008; Cantone and Fisher 2013; Lister et al. 2013). DNA methylation can modulate gene expression in response to exposure to various substances, such as nutrients, smoking, and environmental chemicals (Sandovici et al. 2011; Feil and Fraga 2012; Lee and Pausova 2013). Furthermore, certain lifetime experiences, stressful events, physical training, and memory formation are associated with DNA methylation (Barres et al. 2012; Szyf and Bick 2013; Halder et al. 2016; Saunderson et al. 2016). All these indicate that DNA methylation can affect health and aging by altering cellular genetic programs in response to various non-genetic factors.

Genomic DNA contains epigenetically modifiable DNA bases, and the most studied is 5-methylcytosine. It contains a methyl group covalently attached to position 5 of cytosine found in CpG dinucleotides. DNA methylation levels and patterns vary in different cell types, and for a sample of cells of a particular type, the DNA methylation level at a specific CpG locus is typically represented by the proportion of the 5-methylcytosine-containing CpGs. Most of the epigenetic studies in the literature have used blood samples, and we restrict our interest to blood in this study.

DNA methylation levels at certain CpG dinucleotides change as (chronological) age increases, as shown by both longitudinal and cross-sectional studies (Bjornsson et al. 2008; Calvanese et al. 2009; Talens et al. 2012). Overall, the genome content of 5-methylcytosine tends to decrease as age increases in the elderly. Bollati et al. quantitated 5-methylcytosine levels of repetitive elements in a cohort of individuals aged between 55 and 92 and found both cross-sectional and longitudinal declines with increasing age (Bollati et al. 2009). Heyn et al. also observed a lower 5-methylcytosine level in a centenarian sample compared with a newborn or 26-year-old individual sample (Heyn et al. 2012). In this study, the 26-year-old sample also showed a decrease compared with the newborn sample. In contrast, Fraga et al. found increased 5-methylcytosine levels in 50-year-old twins compared with 3-year-old twins (Fraga et al. 2005). Thus, the age-associated hypomethylation seems contingent on advanced ages (probably past middle age). Interestingly, hypomethylation is also associated with age-related diseases, such as cancer and Alzheimer's disease, which are prevalent among the elderly (Wilson et al. 2007; Pogribny and Beland 2009).

The age-related hypomethylation is delayed in offspring of long-lived parents compared with those of short-lived parents (Gentilini et al. 2013). Because offspring of long-lived parents tends to be healthier than those of short-lived parents (Gentilini et al. 2013; Kim et al. 2013), the delay in loss of DNA methylation in the former could be one of the genome-wide events that go along with healthy aging. Locally, however, there are many genomic sites or regions that gain DNA methylation over time (Fraga et al. 2005; Christensen et al. 2009; Rakyan et al. 2010; Johansson et al. 2013; Jones et al. 2015; Marttila et al. 2015; Sziráki et al. 2018). Moreover, the bulk of the human genome consists of highly methylated repetitive elements (Jones et al. 2015), so age-related hypomethylation of the repetitive elements may well be the major contributor to the global hypomethylation described above. Thus, the association of global loss of DNA methylation with aging and age-related pathology is interesting, but it is likely not applicable to all genomic locations.

As DNA methylation levels at individual CpG sites change over time, DNA methylation patterns of multiple, neighboring CpG sites are likely to change over time as well. Thus, DNA methylation patterns reflect cumulative changes at individual sites. Age-dependent divergence in DNA methylation patterns is evident in MZ twins. Twins and co-twins in MZ pairs increasingly diverge in DNA methylation patterns and gene expression levels, indicating that potentially functional epigenetic changes accumulate over time, likely in response to exposure to non-genetic factors during the life course (Fraga et al. 2005; Talens et al. 2012; Wong et al. 2014).

There are many CpG sites where DNA methylation levels change with age (up to 56,579 CpG sites according to a cross-sectional analysis) (Moore et al. 2016). However, only small portions of them are associated with functional phenotypes, including blood pressure, lung function, hand grip strength, several blood metabolic markers, cognitive functioning, or even mortality (Bell et al. 2012; Hannum et al. 2013; Moore et al. 2016;

Starnawska et al. 2017; Svane et al. 2018). Moreover, DNA methylation at certain CpG sites within the rDNA gene promoter are associated not with age but with cognitive functioning and survival (D'Aquila et al. 2017), indicating that age-related changes alone may not be sufficient to explain the biology of aging. The "epigenetic clock" or similar derivatives of calendar age are based on DNA methylation levels of subsets of age-associated CpG sites (Hannum et al. 2013; Horvath 2013; Weidner and Wagner 2014). However, relevance of these measures to biological aging is questionable. For example, the DNA methylation age measures calculated by Horvath using 353 age-associated CpG sites are comparable to chronological age in mortality prediction (Horvath 2013), but in the presence of age as a covariate, they are no longer significant predictors of mortality (Kim et al. 2017). These observations emphasize the importance of using reliable biological or functional measures of biological age.

Various measures of biological age have been proposed (Kim and Jazwinski 2015), and we have been using the frailty index. Different frailty indexes show similar quantitative and statistical properties if the total numbers of health variables used are statistically valid ($\geq \sim 20$) and the health variables cover diverse health domains (Rockwood et al. 2007; Searle et al. 2008). Frailty indexes have been extensively studied and well characterized. For example, FI_{34} is a significant predictor of mortality (Kim et al. 2013; Kim et al. 2017), and it is associated with various genetic factors and physiological processes, including gut dysbiosis (Kim et al. 2013; Kim et al. 2015; Jazwinski and Kim 2017; Maffei et al. 2017). All these associations were found after adjustment for covariates including age.

In this article, our goal was to find genomic sites or regions in blood samples where DNA methylation correlates with biological age. We calculated a frailty index as a measure of biological age for elderly twins, obtained their genome-wide DNA methylation values, and analyzed the health and epigenome data using association and gene enrichment methods.

Materials and methods

Study subjects

A total of 346 twin subjects were enrolled in the study (Table S1). Their demographic and medical history survey data were collected through the University of Washington Twin Registry (UWTR) and the Mid-Atlantic Twin Registry (MATR). Ages of participants were verified using both documentary evidence (copies of driver's licenses or passports) and demographic questionnaires. All participants provided informed consent prior to enrollment.

Frailty index

A frailty index, called FI_{22}, was constructed, as described previously (Searle et al. 2008; Kim et al. 2013). FI_{22} consists of 22 health and function variables (Table S2). In brief, the health deficit data are quantitative measures or categorical responses from medical history questionnaires. Binary categorical responses were numerically coded, 0 for the absence of the deficit and 1 for the presence of the deficit. Quantitative data and multi-categorical responses were re-coded as shown in the table.

The heritability of FI_{22} was estimated using the equation $h^2 = 2(r_{mz} - r_{dz})$, in which h^2 is the narrow-sense heritability, and r_{mz} and r_{dz} are correlation coefficients of FI_{22} in MZ and DZ twins, respectively (Bell and Spector 2011).

DNA methylation data

DNA samples of participants were prepared from blood spots on FTA cards (Whatman). Leukocyte DNA samples in sufficient quantity (> 500 ng) were subjected to bisulfite conversion per instructions (Zymo Research). DNA methylation measurements of more than 485,000 methylation CpG sites were collected using the Infinium HumanMethylation450 BeadChip assay kit (Illumina) and the Illumina iScan array scanner (University of Utah Genomics Core Facility). The quality of the scanned data was evaluated using not only the internal measures embedded in the BeadChip kit but also functions provided by R packages shinyMethyl (Fortin et al. 2014), minfi (Aryee et al. 2014), and RnBeads (Assenov et al. 2014). Sex was confirmed using the getSex function of minfi. Data that failed to pass this initial QC step were discarded. These initial quality assessments found DNA methylation data from a total of 134 twin participants suitable for further quality evaluation.

In addition to the qualitative assessments of data quality, we took a series of quantitative probe filtering

steps using minfi, DMRcate (Peters et al. 2015), and RnBeads:

1. Excluded CpG sites whose calculated DNA methylation levels across the samples are statistically indistinguishable from the background levels
2. Excluded CpG sites that failed in more than 5% of the samples
3. Excluded six samples that failed to pass the default badSampleCutoff value (10.5) using the getQC function in the minfi package
4. Excluded CpG sites whose beta values have low variation (lower than 0.01)
5. Excluded CpG sites on sex chromosomes
6. Excluded CpG sites within three nucleotides from SNPs
7. Excluded CpG sites interrogated by probes that are known to be "cross-reactive." Cross-reactive probes tend to hybridize to multiple genomic sites (Chen et al. 2013).

Completion of all the quality assessment measures yielded 128 twin samples for further analysis (Table 1), each with DNA methylation data at 409,291 CpG dinucleotides. The scanned raw methylation (m) and unmethylation (u) signals were converted to beta values ($\beta = m/[m + u + 100]$), with offset set at 100, using the minfi package, and the beta values were used for association of DNA methylation levels with the frailty index scores. Density plots of beta values grouped by zygosity, sex, age groups, and frailty index groups are shown in Fig. S1.

Background correction and normalization of DNA methylation data

For background correction, NOOB was applied (Triche Jr. et al. 2013), and for normalization of the background corrected datasets, the SWAN method was applied (Maksimovic et al. 2012). The processed DNA methylation values from whole blood samples were adjusted for various leukocyte compositions using minfi and FlowSorted.Blood.450 k packages according to the regression calibration algorithm (Houseman et al. 2012; Jaffe and Irizarry 2014). The datasets adjusted for cellular heterogeneity were further corrected for batch effects using the ComBat function of the sva package (Leek et al. 2012).

Table 1 Gender, zygosity, and age range of dizygotic (DZ) and monozygotic (MZ) twins subjected to data analysis

Zygosity	Female	Male	Total
DZ	36 (65–87)	22 (65–90)	58 (65–90)
MZ	47 (68–91)	23 (66–90)	70 (66–91)
Total	83 (65–91)	45 (65–90)	128 (65–91)

Age is in parentheses

Measurement of concordance of DNA methylation between co-twins by intraclass correlation (ICC)

Assigning each twin pair in a family as a group, we calculated site-specific ICC values over 10,000 unique methylation sites according to the one-way consistency model using the irr package. The ICC values were compared between MZ and DZ twin pairs using the Wilcoxon rank sum test. As a control, we shuffled twins as if they were unrelated and calculated ICC values over the same methylation sites.

Association analyses

Normality of FI_{22} was tested using the Anderson-Darling and Shapiro-Wilk tests in the r package nortest. Its skewness was tested using the D'Agostino test in the same package. Treating the frailty index as a continuous variable, we used the dmpFinder function of the minfi package and the lmFit function of limma (Ritchie et al. 2015) to seek individual CpG sites associated with FI_{22} after adjustment for sex and age. Also, to take the familial relatedness of twins into account, we estimated the correlation between co-twins using the "duplicateCorrelation" function and included the estimates in the multiple linear models using lmFit. Results from the limma package were repeated using the lme4 package in R, in which the relatedness was included as a random effect term. To find CpG regions consisting of ≥ 2 adjacent CpG sites that are associated with FI_{22}, we used the bumphunter (Jaffe et al. 2012) implemented in minfi, with the default setting. The association of individual CpG sites within each region with FI_{22} was reanalyzed using linear mixed effect modeling, in which FI_{22}, age, and sex were set as fixed effects and family as a random effect.

Ontology analysis

Gene enrichment analysis of a group of selected DNA methylation probes using the Gene Ontology terms consists of three sequential parts: (a) removal of correlated probes, (b) unbiased non-parametric selection of important probes using a random forest method, and (c) GO analysis. The first part is to eliminate DNA methylation probes that are not important but score importance due to their spurious correlation with truly important probes. The cforest permutation importance measure (varimp; see below) provides the "conditional" option, which uses a conditional permutation scheme taking the partial correlation of probes into account. However, we found that application of the varimp function to our datasets, with the conditional option turned on, was extremely time-consuming; therefore, we decided to manually remove inter-correlated probes (one from each pair).

(a) Removal of correlated probes

1. The 409,291 probes were divided into 28 groups (27 × 15,000 probes + the remaining 4291 probes). The main reason for dividing them into smaller groups is to stay within the computing power.
2. The variable age was added to each group, and the findCorrelation function of the caret package was applied to each group to remove probes from highly correlated probe pairs. The function calculates all pair-wise correlation coefficients between probes, and if two probes are highly correlated ($r > 0.6$), the mean absolute correlation of each probe is calculated and the probe with the largest mean absolute correlation is removed. The reason for removing highly correlated probes is that their presence, especially if the correlation is spurious, may yield misleading probe importance outcomes later.
3. The first round of removal of correlated probes above resulted in 203,056 probes in total. These probes were divided into 14 groups (13 × 15,000 + the remaining 8056), and after adding age to each group, the findCorrelation function was applied to each group in the same way as above.
4. The second round of removal of correlated probes above resulted in 102,432 probes in total. These probes were divided into seven groups (6 × 15,000 + the remaining 12,432), and after adding age to each group, the findCorrelation function was applied in the same way.
5. The third round of removal of correlated probes resulted in 96,757 probes remaining in total.

(b) Probe selection

1. The 96,757 probes (+ age) were subjected to recursive partitioning using the cforest function of the party package (Hothorn et al. 2006; Strobl et al. 2007; Strobl et al. 2008) with the default setting (control = cforest_unbiased).
2. The varimp function was applied to the resulting cforest object from each of the seven groups, and the probes whose variable importance was higher than that of age were selected. Application of this step to all seven groups resulted in 4532 DNA methylation probes. Accordingly, DNA methylation profiles of these probes are in low correlation not only with each other but also with age. Furthermore, the importance of these probes in association with FI_{22} is higher than that of age.
3. Distribution of 4532 probes in the genome relative to genes or CpG islands (Fig. S2) does not differ significantly from those of the original probe set ($p \gg 0.1$, $n = 485{,}512$).

(c) GO analysis

1. The 4532 DNA methylation probes were annotated using the getAnnotation function of the minfi package
2. The gometh function of the missMethyl package (Phipson et al. 2016) was applied to the annotated probes, with prior.prob = TRUE to factor the numbers of probes in each gene into the probability calculation. This function maps the probes to Entrez Gene IDs and seeks enriched GO terms using a hypergeometric test.

Results

Heritability of FI_{22}

Because subjects were recruited from two different regions, we first checked whether health data sets used to calculate FI_{22} from these two locations are statistically

homogeneous. The parametric Bartlett's test (bartlett.test) and the non-parametric Fligner-Killeen test (fligner.test) for homogeneity of variance (both in stats package) indicate that the variances of the two sample distributions do not differ ($p \gg 0.05$). Furthermore, the two-sample non-parametric Wilcoxon test indicates that the means of the two sample distributions do not differ either ($p \gg 0.05$). Therefore, the two sets of twin health data were pooled for all subsequent analyses.

FI_{22} was significantly correlated with age ($r = 0.16$, $p = 0.0034$; Fig. 1a), and the narrow-sense heritability was estimated to be 0.27 (95% CI = 0.070~0.46). However, calculation of intraclass correlation of DNA methylation levels between twin-cotwin pairs found no statistical evidence of higher ICC between MZ pairs than between DZ pairs ($p \gg 0.05$).

Single CpG sites associated with FI_{22}

We looked for CpG sites where DNA methylation levels are associated with FI_{22} in MZ or DZ DNA methylation datasets. Using MZ data unadjusted for batch effects (but adjusted for leukocyte heterogeneity), we found 1.2% of the sites analyzed (4919 out of 409,291) showing a FDR q value lower than 0.1 (13 shown in Table 2). Using data adjusted for batch effects, however, we found none with the q values lower than 0.1. Only four sites showed the lowest q value of 0.486 (Table 3). These results demonstrate the striking effect of correcting technical variations and shows the importance of adjusting data for all the known confounders before downstream analysis. This is especially true with large-scale, high-throughput data. With additional adjustment for age, sex, and within-pair correlation, we found 40 sites with the q value between 0.2 and 0.5 (five shown in Table 4). Using DZ twin data, adjusted or unadjusted, we found no sites with the q values lower than 0.5.

Multiple CpG sites associated with FI_{22}

Using the batch-adjusted data of MZ twins, we looked for genomic regions containing ≥ 2 adjacent CpG sites that are associated with FI_{22} scores and found two regions with family-wise error rates below 0.5 (Table 5). These are a 450-bp region on chromosome 17 and a 34-bp region on chromosome 5, the latter of which includes cg17588578 (row 2 in Table 2). Individual CpG sites within the two regions were reanalyzed using linear mixed effect modeling, in which FI_{22}, age, and sex were set as fixed effects, and the family ID was set as a random effect. Each of the three CpG sites in the chromosome 5 region showed a significant correlation with DNA methylation level ($p < 0.01$), whereas none of the CpG sites within the chromosome 17 region did (Fig. 2). Note that age was not significantly associated with DNA methylation for all three chromosome 5 probes, whereas the chromosome 17 probe shown in Fig. 2 was. Thus, although FI_{22} is correlated with age, association of DNA methylation with FI_{22} does not always go together with its association with age.

The chromosome 5 region containing the three CpG sites is located at the 5′ side of *PCDHGA3* (Protocadherin Gamma Subfamily A, 3) (Fig. 3), within tandemly linked protocadherin gamma gene clusters (Supplementary Fig. S3). It encodes a protein for homophilic cell-cell adhesion (Weiner and Jontes 2013). The CpG region falls on the 5′ end of a CpG island, histone marks, a DNase I sensitivity site, and binding sites for CCCTC-binding factor (CTCF) and a RAD21 cohesion complex component. The CTCF-binding site is also supported by ChromHMM tracks, which display functional chromatin states for different cell types (Ernst et al. 2011; Ernst and Kellis 2012). CTCF-binding sites frequently overlap with RAD21-binding sites (Nitzsche et al. 2011; Monahan et al. 2012).

Gene ontology analysis

The association methods described so far are based on parametric regression that will work best when the dependent variable follows a normal distribution. FI_{22} is not normally distributed ($p \ll 0.01$; Fig. 1b) and right-skewed ($p \ll 0.01$), which is typical for frailty indexes (Mitnitski et al. 2001). We also have the "small n large v" problem (the subject number n is greatly exceeded by the variable number v), which also discourages using parametric methods (Strobl et al. 2009). Therefore, we used a Random Forest approach and compared the outcomes. The Random forest method, which is based on non-parametric recursive partitioning, can handle large numbers of variables and provide variable importance scores for individual variables (Strobl et al. 2009).

Using the Random Forest, we collected CpG sites whose variable importance is higher than that of age and inferred enriched biological functions using the GO Consortium database. Thus, first, we removed CpG sites

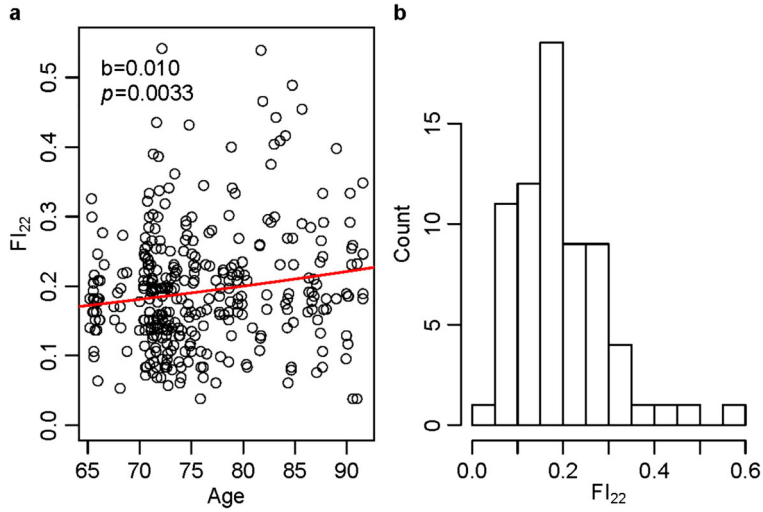

Fig. 1 A scatter plot of FI_{22} scores by age (**a**) and a histogram of counts of subjects with the indicated FI_{22} score brackets (**b**). **a** b is the coefficient estimate of the exponential function ($e^{(b*age)}$) and its p value under the null hypothesis that the slope = 0. FI_{22} and age are also linearly correlated, with $r = 0.01$ and $p = 0.0033$

that were highly correlated ($r > 0.6$) not only with other CpG sites but also with age. This procedure is to prevent probes of low importance from being selected due to their spurious correlation with true probes of high importance. Second, age and CpG sites were recursively partitioned such that each partition contained CpG variables with similar predictability on FI_{22}, and CpG sites that were higher in variable importance than age were selected. Third, the selected sites were tested for GO term enrichment.

The GO analysis found 68 enriched GO terms from the 4532 selected probes: 53 biological processes (BP), seven cellular components (CC), and eight molecular functions (MF) (Table S3). Of these, the most significant term is "homophilic cell adhesion via plasma membrane adhesion molecules" (GO:0007156 at the top in Fig. 4; $p = 5.46e{-}07$, $q = 0.011$). According to the Ancestor chart provided by the QuickGo browser (Binns et al. 2009), there are four additional terms related to GO:0007156 (Fig. 5), and all these terms were included in the enriched group, with q values ranging from 0.043 to 0.20 (Fig. 4). Also, included in the enrichment group are cell-cell signaling and calcium-ion binding, which are often annotated to the protocadherin proteins as is GO:0007156 ("co-occurring" terms, with a superscript asterisk, in Fig. 4). Thus, these results indicate that the non-parametrically selected DNA methylation probes are enriched with probes associated with genes involved in homophilic cell-cell adhesion. According to the Gene Ontology Annotation (GOA) database (Barrell et al. 2009), homophilic cell adhesion via plasma membrane adhesion molecules (GO: 0007156) and cell adhesion (GO: 0007155) are GO biological process terms

Table 2 CpG sites associated with FI_{22}, using DNA methylation data from MZ twins, adjusted for leukocyte heterogeneity only. Only 13 sites with the lowest q values are shown

Row	Probe	Coefficient	t value	p	q
1	cg20810993	−0.370	−5.866	1.51E−07	0.046
2	cg17588578	0.894	5.655	3.49E−07	0.053
3	cg15709766	0.633	5.343	1.18E−06	0.070
4	cg12867728	0.460	5.340	1.19E−06	0.070
5	cg16421157	0.702	5.294	1.43E−06	0.070
6	cg18168101	0.243	5.268	1.58E−06	0.070
7	cg03340356	0.650	5.264	1.60E−06	0.070
8	cg14074052	−0.110	−5.226	1.85E−06	0.071
9	cg21858380	0.328	5.158	2.40E−06	0.074
10	cg13365524	0.427	5.155	2.43E−06	0.074
11	cg10502121	0.385	5.089	3.12E−06	0.076
12	cg00861009	0.297	5.054	3.57E−06	0.076
13	cg13072214	0.219	−4.279	6.22E−05	0.076

Table 3 CpG sites associated with FI_{22}, using DNA methylation data from MZ twins, adjusted for both leukocyte heterogeneity and batch difference

Row*	Probe	Coefficient	t value	p	q
3	cg15709766	0.612	5.267	1.58E−06	0.486
12	cg00861009	0.240	5.082	3.21E−06	0.486
2	cg17588578	0.739	4.989	4.57E−06	0.486
5	cg16421157	0.636	4.951	5.27E−06	0.486

*Row number in Table 2

Table 4 Individual CpG sites associated with FI_{22}, using DNA methylation data from MZ twins, adjusted for leukocyte heterogeneity, differences in batch, age, sex, and within-pair correlation. Only five out of 40 sites with q value lower than 0.5 are shown

Row*	Probe	Chromosome	logFC	t value	p	q
2	cg17588578	5	5.76	5.530	5.33E−07	0.22
5	cg16421157	6	4.69	5.233	1.70E−06	0.35
13	cg13072214	8	−4.84	−5.031	3.70E−06	0.39
3	cg15709766	19	4.35	5.024	3.80E−06	0.39
12	cg00861009	2	2.49	4.65	1.55E−05	0.47

*Row number in Table 2. logFC represents $\log_2$ (adjusted average change in DNA methylation for a unit change in FI_{22})

assigned to PCDHG3A. The results shown in Table S3 and Fig. 4 are after adjustment for the number of probes assigned to each gene. Without the adjustment in the probability calculation, different results were obtained with much lower p and q values (Supplementary Table S4).

Another noticeable group of terms involves the G protein-coupled receptor signaling pathway (3 BPs and 1 MF), with q values from 0.043 to 0.087 (Fig. 4). Terms closely related to GO:0007188 include various terms in categories of cellular signaling and response to stimulus.

Discussion

Our results indicate that DNA methylation levels at CpG sites linked to *PCDHGA3* are associated with FI_{22}, and this association can be extended to other members of clustered protocadherin genes, based on previous studies (see below). The first set of results, generated by taking the parametric regression approach, points to DNA methylation at three CpG sites in the promoter region of *PCDHGA3*. The second set obtained independently by taking a non-parametric machine learning approach followed by GO ontology analysis also indicates that DNA methylation sites linked to genes involved in cell-cell adhesion are enriched among CpG sites whose DNA methylation levels are predictive of FI_{22}. Although interpretation of random forest outputs could be controversial, as is that of any other non-parametric method, we found two different approaches pointing in the same direction.

PCDHGA3 is a member of the gamma gene (*PCDHG*) cluster located on chromosome 5q31, along with the alpha (*PCDHA*) and beta (*PCDHB*) gene clusters. Highly conserved among vertebrates, these clustered genes are organized in tandem like the immunoglobin genes, and matched combinations of the expressed isoforms are required for fully functional neurodevelopment (Chen and Maniatis 2013; Hasegawa et al. 2017). The gamma isoforms are necessary for neuronal complexity by promoting dendrite arborization through homophilic cell-cell interactions (Schreiner and Weiner 2010; Molumby et al. 2016).

Several CpG sites in clustered protocadherin genes have appeared in various lists in which DNA methylation levels are associated with age in blood samples (Rakyan et al. 2010; Bell et al. 2012; Salpea et al. 2012; McClay et al. 2014). Hints of the association of DNA methylation at protocadherin genes with biological age were given by Hannum et al. 2013 and Slieker et al. 2016. The former group selected CpG sites, including a CpG site in *PCDHB1*, based on a predictive model of age with adjustment for BMI and diabetes. Using DNA methylation data from more than 3000 individuals, the latter group selected CpG sites where

Table 5 Regions of CpG sites associated with FI_{22}, using DNA methylation data from MZ twins, adjusted for leukocyte heterogeneity, differences in batch, age, sex, and within-pair correlation

chr	Start	End	bp	Value[1]	n[2]	p area[3]	P area[4]
17	6,558,365	6,558,815	450	−0.80	3	0.0083	0.23
5	140,723,549	140,723,583	34	0.67	3	0.016	0.38

1, the average of the estimated coefficients; 2, number of probes in the region; 3, based on the area under the bump; 4, adjusted for multiple comparison by permutation-based family-wise error rate calculation

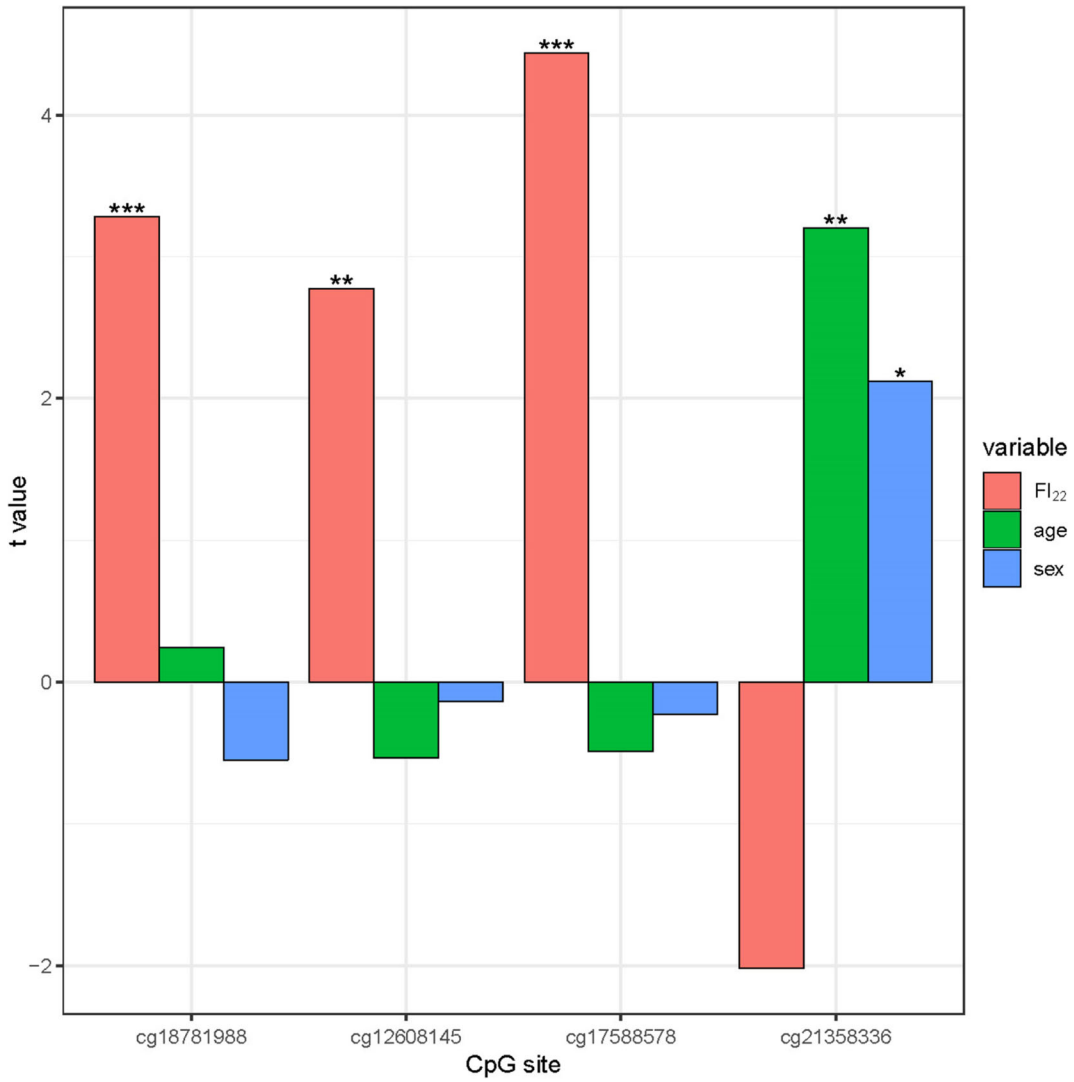

Fig. 2 A bar plot of effect sizes (t values = coefficient estimates/standard errors) from linear mixed effects models of FI_{22} regressed on beta values of indicated CpG sites, with age and sex as fixed covariates and family ID as a random effect. At the first three sites on chromosome 5 (x-axis), only DNA methylation levels are significantly associated with FI_{22}. The fourth site is one of the probes on chromosome 17 (Table 5), where DNA methylation levels are significantly associated with age, but not with FI_{22} ($p = 0.508$). * is for $0.01 < p \leq 0.05$, ** $0.001 < p \leq 0.01$, *** $p \leq 0.001$

variability in DNA methylation increases with age, and these "age-related variably methylated positions (aVMP)" include many CpG sites across all three protocadherin gene clusters on chromosome 5. However, these aVMPs are yet to be tested for their association with biological age. The ultimate test for functionality of a CpG site can be shown by its association with mortality, and this functionality of the CpG sites linked to *PCDHGA3* is yet to be tested. However, a recent study suggests that they, especially cg17588578, are likely to be associated with mortality. Svane et al. found more than 2800 CpG sites associated individually with mortality, and this list contains a CpG site linked to *PCDHGA3* (Svane et al. 2018). Recently, using a "phenotypic age" constructed from nine lab blood test measurements and age, Levine et al. compiled more than 500 CpG sites (DNAm PhenoAge) that are collectively predictive of phenotypic age, a risk factor for mortality, and other health phenotypes (Levine et al. 2018). This list of 500 sites contains several protocadherin genes (*PCDHB* and *PCDHG*). A drawback of DNAm PhenoAge is its very heavy reliance on

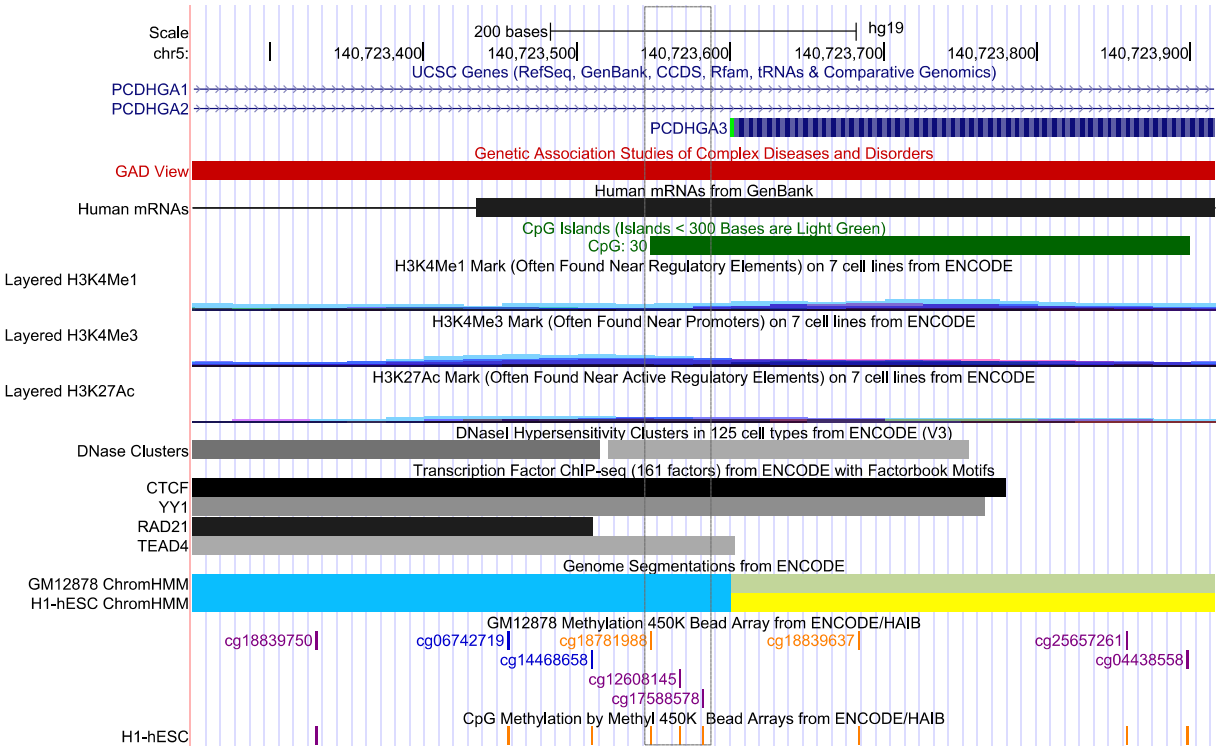

Fig. 3 Genome features around cg18781988, cg12608145, and cg17588578 (enclosed in a rectangle) on chromosome 5, shown by the UCSC Genome Browser (hg19; GRCh37). The region containing the three CpG sites overlaps a stretch marked by Genetic Association Studies of Complex Diseases and Disorders (red), a CpG Island (green), histone marks, a DNase I hypersensitivity site, transcription factor-binding sites (especially, CTCF/cohesion binding sites), and a CCCTC-binding factor (CTCF)-enriched chromatin segment (blue). Colors in the histone marks represent results from different cell lines, and the low peak levels, which correspond to low enrichment levels of histone marks as determined by ChIP-seq assays, predicts a minor, if any, regulatory role of the histone marks in this region. The DNase Clusters track shows DNase I hypersensitive sites with the darkness proportional to sensitivity. The Transcription Factor ChIP-seq track shows transcription factor-binding sites based on the ENCODE ChIP-seq experiments, and the darkness is proportional to the signal strength. The ChromHMM tracks show chromatin segments of different functional states (Ernst et al. 2011): the light blue color represents an insulator function and the light green- or yellow-colored segment is a region with weak enhancer or weak transcription activity. Colors of DNA methylation probe names represent methylation status in a given cell line: orange, methylated; purple, partially methylated; blue, unmethylated

chronological age for predictive value. Our study using a measure of biological age, FI_{22}, which does not rely on chronological age, strongly indicates that indeed, protocadherin DNA methylation on chromosome 5, especially *PCDHGA3*, is associated with healthy aging.

Regression outputs for the three *PCDHGA3* CpG sites indicate that an increase in DNA methylation, especially at cg17588578, is positively associated with an increase in FI_{22} (Table 6). In other words, hypermethylation at these sites seems to correlate with unhealthy aging. Normal expression of this gene is required for proper development and cognitive and physical functioning. If we adopt the generalization that hypermethylation corresponds to lower gene expression (Martinowich et al. 2003; Calvanese et al. 2009; Straussman et al. 2009), we can tentatively conclude that reduced expression of *PCDHGA3,* or possibly any of the other related, clustered protocadherin genes, by increased DNA methylation of its promoter leads to increased FI_{22}, i.e., higher risk for unhealthy aging.

Data available so far, mostly from murine studies, accord with our tentative conclusion. Various protocadherin isoforms are differentially expressed during embryo development, and CTCF/cohesin-binding sites found in many protocadherin gene promoters are important in differential regulation (Chen and Maniatis 2013; Weiner and Jontes 2013; Hirayama and Yagi 2017). CTCF binding can either facilitate or block enhancer function, depending on whether an enhancer and its target gene are included in the same chromatin loop

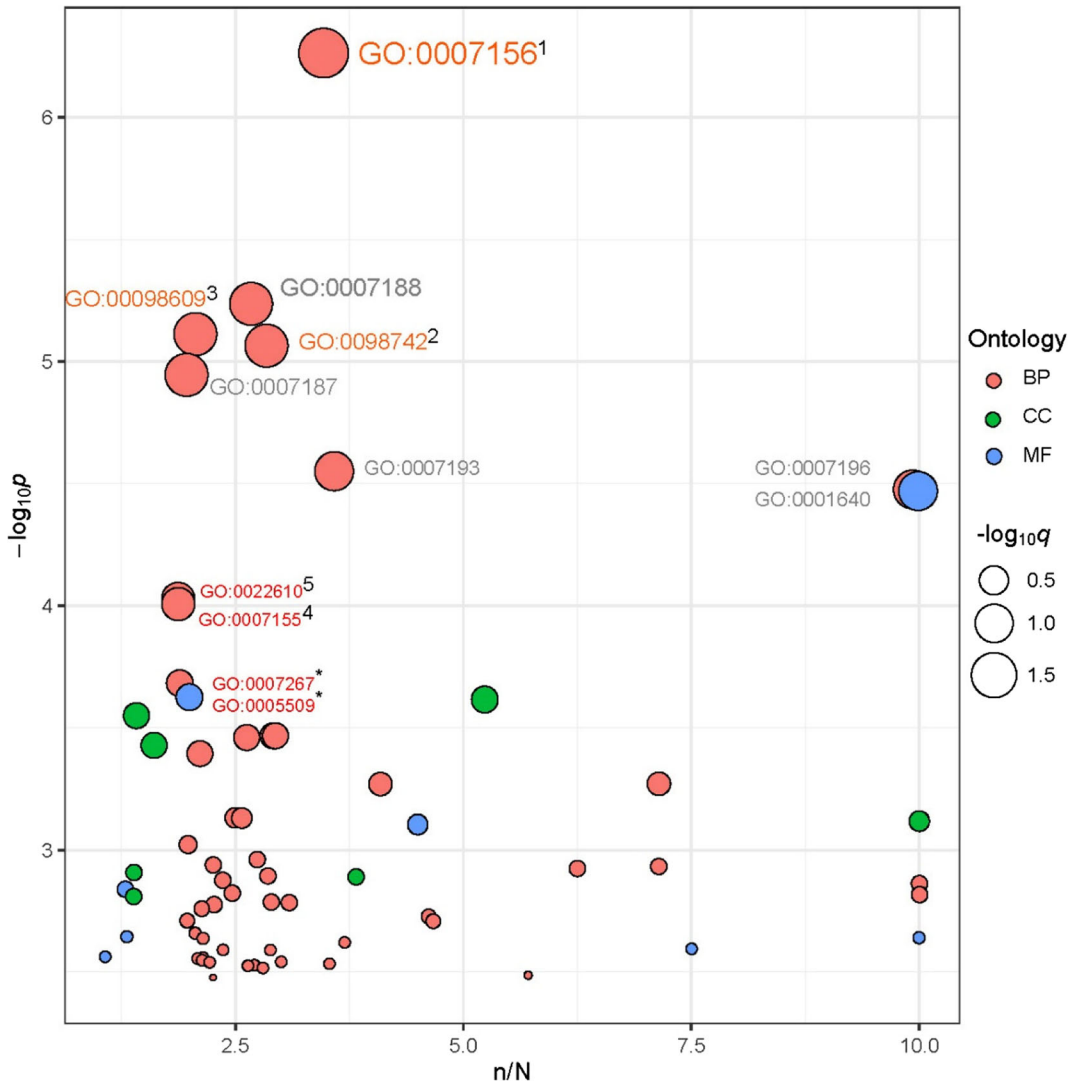

Fig. 4 A bubble plot of 68 GO terms enriched among identifiable genes associated with 4532 DNA methylation probes selected by the random forest method (Table S3). *n*, number of genes differentially methylated; *N*, number of genes in the GO; *p*, *p* value for over-representation of the GO term; *q*, false discovery rate; BP, biological process; CC, cell component; MF, molecular function. GO IDs shown in red are for terms related to homophilic cell adhesion, and those shown in gray are for terms related to G protein-coupled receptor pathways. The superscript numbers correspond to those in Fig. 5, and the two GO terms with a superscript asterisk are co-occurring terms with GO:0007156 (GO:007267, cell-cell adhesion; GO:0005509, calcium ion binding)

or separated into different loops (Ong and Corces 2014). There are two CpG sites in the consensus CTCF-binding sequence, CCGCGNGGNGGCAG (Kim et al. 2007), and CTCF binding to the sequence can be affected by DNA methylation (Bell and Felsenfeld 2000; Maurano et al. 2012; Wang et al. 2012). In addition, the promoter regions of many clustered protocadherin genes are differentially methylated, adding another layer of complexity (Toyoda et al. 2014). DNA methylation levels at the 5′ regions of clustered *PCDHA* genes inversely correlate with gene expression levels, and 5-azacytidine treatment induces expression of silent *PCDHA* genes (Kawaguchi et al. 2008). Hypermethylation is also associated with reduced expression of *PCDHB* and *PCDHG* genes (Dallosso et al. 2009), and reduced or deregulated protocadherin expression is associated with various neurological, psychiatric, and cardiovascular disorders, and tumor growth (Dallosso et al. 2009; Ortega et al. 2016; Zhang et al. 2016; El Hajj et al. 2017). Interestingly, hypermethylation at the promoter regions of the

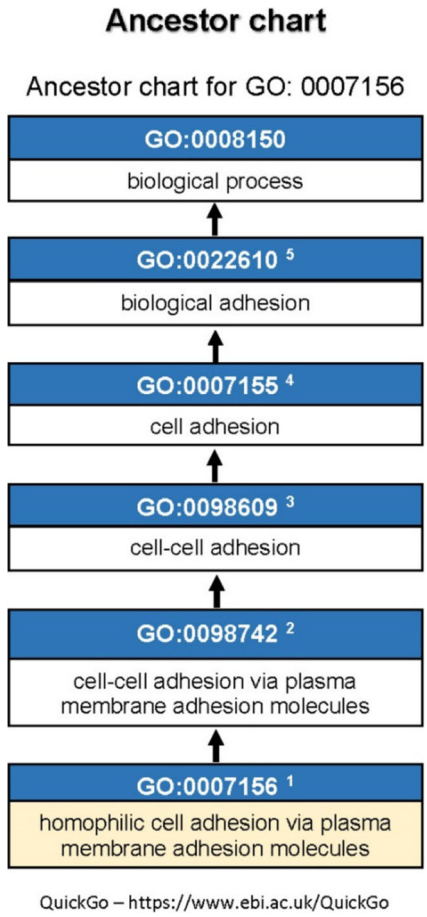

Fig. 5 The "Ancestor chart" of QuickGo lists parent GO terms of a query term at the bottom. The superscript numbers correspond to those in Fig. 4

clustered protocadherin genes is associated with histories of childhood abuse and disadvantaged socioeconomic status in adult humans (Borghol et al. 2012; Suderman et al. 2012), and a history of low maternal care in adult rats (Suderman et al. 2012). These observations are consistent with the notion that an increase in DNA methylation level at the 5′ side of *PCDHGA3* can diminish its expression, increasing the risk of poor health and aging, and this notion can be easily extended to other related clustered protocadherin genes on chromosome 5.

Unlike other studies, our study of elderly twins did not find higher intraclass correlation with MZ twins than with DZ twins. This implies no significant heritability of DNA methylation. Heritability of DNA methylation varies with loci examined (Kaminsky et al. 2009), with some regions showing negligible heritability (Gervin et al. 2011). No difference in ICC between MZ and DZ may indicate dominance of non-genetic or stochastic DNA methylation changes over genetic differences. One such non-genetic factor could be the so-called "epigenetic drift" as a function of age. The twins in our study were all above 65 and into their 90s.

In sum, using health and DNA methylation data from elderly twins, we found that DNA methylation at CpG sites in the promoter region of *PCDHGA3* is associated with FI_{22}, a measure of biological age. *PCDHGA3* belongs to the protocadherin gene clusters all located in a single locus, and DNA methylation levels at CpG sites linked to several other protocadherin genes have been associated with age or several phenotypic measures. Data available in the literature suggest that dysregulated expression of protocadherin genes is associated with various pathologies and DNA methylation is a key modifier of expression of the genes. Furthermore, DNA methylation patterns in the clustered protocadherin genes have been associated with disadvantaged lifetime experiences. All these observations suggest that DNA methylation at the protocadherin genes is likely to be a key mediator of healthy aging. The methylation status at this locus is also a measure of biological age and thus of mortality hazard.

Table 6 Association of the three CpG sites, within the chromosome 5 region, with FI_{22}, using DNA methylation data from MZ twins, adjusted for leukocyte heterogeneity, differences in batch, age, sex, and within-pair correlation

Probe	Position	*b*	se (*b*)	*p*
cg18781988	140,723,549	4.80	1.48	0.0019
cg12608145	140,723,568	3.93	1.40	0.0066
cg17588578	140,723,583	5.42	1.04	2.17e−06

Acknowledgments We thank participants in our studies. We also thank the Mid-Atlantic Twin Registry at Virginia Commonwealth University and the University of Washington Twin Registry for recruiting twins and collecting health data.

Funding information This study was supported by the National Institute of General Medical Sciences of the National Institutes of Health (P20GM103629) to S.M.J. and S.K.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30: 1363–1369. https://doi.org/10.1093/bioinformatics/btu049

Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C (2014) Comprehensive analysis of DNA methylation data with RnBeads. Nat Methods 11:1138–1140. https://doi.org/10.1038/nmeth.3115

Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R (2009) The GOA database in 2009–an integrated Gene Ontology Annotation resource. Nucleic Acids Res 37: D396–D403. https://doi.org/10.1093/nar/gkn803

Barres R et al (2012) Acute exercise remodels promoter methylation in human skeletal muscle. Cell Metab 15:405–411. https://doi.org/10.1016/j.cmet.2012.01.001

Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405:482–485. https://doi.org/10.1038/35013100

Bell JT, Spector TD (2011) A twin approach to unraveling epigenetics. Trends Genet 27:116–125. https://doi.org/10.1016/j.tig.2010.12.005

Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin SY, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, Small KS, The MuTHER Consortium, Dermitzakis ET, McCarthy MI, Mill J, Spector TD, Deloukas P (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet 8:e1002629. https://doi.org/10.1371/journal.pgen.1002629

Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R (2009) QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 25:3045–3046. https://doi.org/10.1093/bioinformatics/btp536

Bjornsson HT et al (2008) Intra-individual change over time in DNA methylation with familial clustering. JAMA 299:2877–2883. https://doi.org/10.1001/jama.299.24.2877

Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P, Baccarelli A (2009) Decline in genomic DNA methylation through aging in a cohort of elderly subjects. Mech Ageing Dev 130:234–239. https://doi.org/10.1016/j.mad.2008.12.003

Borghol N, Suderman M, McArdle W, Racine A, Hallett M, Pembrey M, Hertzman C, Power C, Szyf M (2012) Associations with early-life socio-economic position in adult DNA methylation. Int J Epidemiol 41:62–74. https://doi.org/10.1093/ije/dyr147

Calvanese V, Lara E, Kahn A, Fraga MF (2009) The role of epigenetics in aging and age-related diseases. Ageing Res Rev 8:268–276. https://doi.org/10.1016/j.arr.2009.03.004

Cantone I, Fisher AG (2013) Epigenetic programming and reprogramming during development. Nat Struct Mol Biol 20:282–289. https://doi.org/10.1038/nsmb.2489

Castillo-Fernandez JE, Spector TD, Bell JT (2014) Epigenetics of discordant monozygotic twins: implications for disease. Genome Med 6:60. https://doi.org/10.1186/s13073-014-0060-z

Chen WV, Maniatis T (2013) Clustered protocadherins. Development 140:3297–3302. https://doi.org/10.1242/dev.090621

Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8:203–209. https://doi.org/10.4161/epi.23470

Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh RF, Wiencke JK, Kelsey KT (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet 5:e1000602. https://doi.org/10.1371/journal.pgen.1000602

D'Aquila P, Montesanto A, Mandalà M, Garasto S, Mari V, Corsonello A, Bellizzi D, Passarino G (2017) Methylation of the ribosomal RNA gene promoter is associated with aging and age-related decline. Aging Cell 16:966–975. https://doi.org/10.1111/acel.12603

Dallosso AR, Hancock AL, Szemes M, Moorwood K, Chilukamarri L, Tsai HH, Sarkar A, Barasch J, Vuononvirta R, Jones C, Pritchard-Jones K, Royer-Pokora B, Lee SB, Owen C, Malik S, Feng Y, Frank M, Ward A, Brown KW, Malik K (2009) Frequent long-range epigenetic silencing of protocadherin gene clusters on chromosome 5q31 in Wilms' tumor. PLoS Genet 5:e1000745. https://doi.org/10.1371/journal.pgen.1000745

El Hajj N, Dittrich M, Haaf T (2017) Epigenetic dysregulation of protocadherins in human disease. Semin Cell Dev Biol 69: 172–182. https://doi.org/10.1016/j.semcdb.2017.07.007

Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. Nat Methods 9:215–216. https://doi.org/10.1038/nmeth.1906

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473:43–49. https://doi.org/10.1038/nature09906

Feil R, Fraga MF (2012) Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 13:97–109. https://doi.org/10.1038/nrg3142

Fortin JP, Fertig E, Hansen K (2014) shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. F1000Res 3:175. https://doi.org/10.12688/f1000research.4680.2

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A

102:10604–10609. https://doi.org/10.1073/pnas.0500398102

Gentilini D, Mari D, Castaldi D, Remondini D, Ogliari G, Ostan R, Bucci L, Sirchia SM, Tabano S, Cavagnini F, Monti D, Franceschi C, di Blasio AM, Vitale G (2013) Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. Age (Dordr) 35:1961–1973. https://doi.org/10.1007/s11357-012-9463-1

Gervin K, Hammero M, Akselsen HE, Moe R, Nygard H, Brandt I, Gjessing HK, Harris JR, Undlien DE, Lyle R (2011) Extensive variation and low heritability of DNA methylation identified in a twin study. Genome Res 21:1813–1821. https://doi.org/10.1101/gr.119685.110

Gudmundsson H, Gudbjartsson DF, Frigge M, Gulcher JR, Stefansson K (2000) Inheritance of human longevity in Iceland. Eur J Hum Genet 8:743–749. https://doi.org/10.1038/sj.ejhg.5200527

Halder R, Hennion M, Vidal RO, Shomroni O, Rahman RU, Rajput A, Centeno TP, van Bebber F, Capece V, Vizcaino JCG, Schuetz AL, Burkhardt S, Benito E, Sala MN, Javan SB, Haass C, Schmid B, Fischer A, Bonn S (2016) DNA methylation changes in plasticity genes accompany the formation and maintenance of memory. Nat Neurosci 19:102–110. https://doi.org/10.1038/nn.4194

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda SV, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T, Zhang K (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49:359–367. https://doi.org/10.1016/j.molcel.2012.10.016

Hasegawa S, Kobayashi H, Kumagai M, Nishimaru H, Tarusawa E, Kanda H, Sanbo M, Yoshimura Y, Hirabayashi M, Hirabayashi T, Yagi T (2017) Clustered protocadherins are required for building functional neural circuits. Front Mol Neurosci 10:114. https://doi.org/10.3389/fnmol.2017.00114

Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW (1996) The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. Hum Genet 97:319–323

Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ, Puca AA, Sayols S, Pujana MA, Serra-Musach J, Iglesias-Platas I, Formiga F, Fernandez AF, Fraga MF, Heath SC, Valencia A, Gut IG, Wang J, Esteller M (2012) Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci U S A 109:10522–10527. https://doi.org/10.1073/pnas.1120658109

Hirayama T, Yagi T (2017) Regulation of clustered protocadherin genes in individual neurons. Semin Cell Dev Biol 69:122–130. https://doi.org/10.1016/j.semcdb.2017.05.026

Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14:R115. https://doi.org/10.1186/gb-2013-14-10-r115

Hothorn T, Buhlmann P, Dudoit S, Molinaro A, van der Laan MJ (2006) Survival ensembles. Biostatistics 7:355–373. https://doi.org/10.1093/biostatistics/kxj011

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86. https://doi.org/10.1186/1471-2105-13-86

Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol 15:R31. https://doi.org/10.1186/gb-2014-15-2-r31

Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA (2012) Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol 41:200–209. https://doi.org/10.1093/ije/dyr238

Jazwinski SM, Kim S (2017) Metabolic and genetic markers of biological age. Front Genet 8:64. https://doi.org/10.3389/fgene.2017.00064

Johansson A, Enroth S, Gyllensten U (2013) Continuous aging of the human DNA methylome throughout the human lifespan. PLoS One 8:e67378. https://doi.org/10.1371/journal.pone.0067378

Jones MJ, Goodman SJ, Kobor MS (2015) DNA methylation and healthy human aging. Aging Cell 14:924–932. https://doi.org/10.1111/acel.12349

Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GHT, Wong AHC, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, McRae AF, Visscher PM, Montgomery GW, Gottesman II, Martin NG, Petronis A (2009) DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet 41:240–245. https://doi.org/10.1038/ng.286

Kawaguchi M, Toyama T, Kaneko R, Hirayama T, Kawamura Y, Yagi T (2008) Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin-alpha gene cluster. J Biol Chem 283:12064–12075. https://doi.org/10.1074/jbc.M709648200

Kerber RA, O'Brien E, Smith KR, Cawthon RM (2001) Familial excess longevity in Utah genealogies. J Gerontol A Biol Sci Med Sci 56:B130–B139

Kim S, Jazwinski SM (2015) Quantitative measures of healthy aging and biological age. Healthy Aging Res 4 https://doi.org/10.12715/har.2015.4.26

Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128:1231–1245. https://doi.org/10.1016/j.cell.2006.12.048

Kim S, Welsh DA, Cherry KE, Myers L, Jazwinski SM (2013) Association of healthy aging with parental longevity. Age (Dordr) 35:1975–1982. https://doi.org/10.1007/s11357-012-9472-0

Kim S, Welsh DA, Myers L, Cherry KE, Wyckoff J, Jazwinski SM (2015) Non-coding genomic regions possessing enhancer and silencer potential are associated with healthy aging and exceptional survival. Oncotarget 6:3600–3612

Kim S, Myers L, Wyckoff J, Cherry KE, Jazwinski SM (2017) The frailty index outperforms DNA methylation age and its derivatives as an indicator of biological age. Geroscience 39:83–92. https://doi.org/10.1007/s11357-017-9960-3

Lee KW, Pausova Z (2013) Cigarette smoking and DNA methylation. Front Genet 4:132. https://doi.org/10.3389/fgene.2013.00132

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28:882–883. https://doi.org/10.1093/bioinformatics/bts034

Levine ME et al. (2018) An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY) 10:573–591 https://doi.org/10.18632/aging.101414

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM, Ecker JR (2013) Global epigenomic reconfiguration during mammalian brain development. Science 341:1237905. https://doi.org/10.1126/science.1237905

Maffei VJ, Kim S, Blanchard E, Luo M, Jazwinski SM, Taylor CM, Welsh DA (2017) Biological aging and the human gut microbiota. J Gerontol A Biol Sci Med Sci 72:1474–1482. https://doi.org/10.1093/gerona/glx042

Maksimovic J, Gordon L, Oshlack A (2012) SWAN: subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol 13:R44. https://doi.org/10.1186/gb-2012-13-6-r44

Martinowich K et al (2003) DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science 302:890–893. https://doi.org/10.1126/science.1090842

Marttila S, Kananen L, Häyrynen S, Jylhävä J, Nevalainen T, Hervonen A, Jylhä M, Nykter M, Hurme M (2015) Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. BMC Genomics 16:179. https://doi.org/10.1186/s12864-015-1381-z

Maurano MT, Wang H, Kutyavin T, Stamatoyannopoulos JA (2012) Widespread site-dependent buffering of human regulatory polymorphism. PLoS Genet 8:e1002599. https://doi.org/10.1371/journal.pgen.1002599

McClay JL et al (2014) A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. Hum Mol Genet 23:1175–1185. https://doi.org/10.1093/hmg/ddt511

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454:766–770. https://doi.org/10.1038/nature07107

Mitchell BD, Hsueh WC, King TM, Pollin TI, Sorkin J, Agarwala R, Schäffer AA, Shuldiner AR (2001) Heritability of life span in the Old Order Amish. Am J Med Genet 102:346–352 https://doi.org/10.1002/ajmg.1483 [pii]

Mitnitski AB, Mogilner AJ, Rockwood K (2001) Accumulation of deficits as a proxy measure of aging. ScientificWorldJournal 1:323–336. https://doi.org/10.1100/tsw.2001.58

Molumby MJ, Keeler AB, Weiner JA (2016) Homophilic protocadherin cell-cell interactions promote dendrite complexity. Cell Rep 15:1037–1050. https://doi.org/10.1016/j.celrep.2016.03.093

Monahan K, Rudnick ND, Kehayova PD, Pauli F, Newberry KM, Myers RM, Maniatis T (2012) Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-alpha gene expression. Proc Natl Acad Sci U S A 109:9125–9130. https://doi.org/10.1073/pnas.1205074109

Moore AZ, Hernandez DG, Tanaka T, Pilling LC, Nalls MA, Bandinelli S, Singleton AB, Ferrucci L (2016) Change in epigenome-wide DNA methylation over 9 years and subsequent mortality: results from the InCHIANTI Study. J Gerontol A Biol Sci Med Sci 71:1029–1035. https://doi.org/10.1093/gerona/glv118

Nitzsche A, Paszkowski-Rogacz M, Matarese F, Janssen-Megens EM, Hubner NC, Schulz H, de Vries I, Ding L, Huebner N, Mann M, Stunnenberg HG, Buchholz F (2011) RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. PLoS One 6:e19470. https://doi.org/10.1371/journal.pone.0019470

Ong C-T, Corces VG (2014) CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet 15:234–246. https://doi.org/10.1038/nrg3663

Ortega A et al. (2016) New cell adhesion molecules in human ischemic cardiomyopathy. PCDHGA3 implications in decreased stroke volume and ventricular dysfunction. PLoS One 11:e0160168 https://doi.org/10.1371/journal.pone.0160168

Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, Clark SJ, Molloy PL (2015) De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin 8:6–16. https://doi.org/10.1186/1756-8935-8-6

Phipson B, Maksimovic J, Oshlack A (2016) missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics 32:286–288. https://doi.org/10.1093/bioinformatics/btv560

Pogribny IP, Beland FA (2009) DNA hypomethylation in the origin and pathogenesis of human diseases. Cell Mol Life Sci 66:2249–2261. https://doi.org/10.1007/s00018-009-0015-5

Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res 20:434–439. https://doi.org/10.1101/gr.103101.109

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47. https://doi.org/10.1093/nar/gkv007

Rockwood K, Andrew M, Mitnitski A (2007) A comparison of two approaches to measuring frailty in elderly people. J Gerontol A Biol Sci Med Sci 62:738–743 doi:62/7/738 [pii]

Salpea P, Russanova VR, Hirai TH, Sourlingas TG, Sekeri-Pataryas KE, Romero R, Epstein J, Howard BH (2012) Postnatal development- and age-related changes in DNA-methylation patterns in the human genome. Nucleic Acids Res 40:6477–6494. https://doi.org/10.1093/nar/gks312

Sandovici I, Smith NH, Nitert MD, Ackers-Johnson M, Uribe-Lewis S, Ito Y, Jones RH, Marquez VE, Cairns W, Tadayyon M, O'Neill LP, Murrell A, Ling C, Constancia M, Ozanne SE (2011) Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets. Proc Natl Acad Sci U S A 108:5449–5454. https://doi.org/10.1073/pnas.1019007108

Saunderson EA, Spiers H, Mifsud KR, Gutierrez-Mecinas M, Trollope AF, Shaikh A, Mill J, Reul JMHM (2016) Stress-induced gene expression and behavior are controlled by DNA methylation and methyl donor availability in the dentate gyrus. Proc Natl Acad Sci U S A 113:4830–4835. https://doi.org/10.1073/pnas.1524857113

Schreiner D, Weiner JA (2010) Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. Proc Natl Acad Sci U S A 107:14893–14898. https://doi.org/10.1073/pnas.1004526107

Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K (2008) A standard procedure for creating a frailty index. BMC Geriatr 8:24. https://doi.org/10.1186/1471-2318-8-24

Slieker RC et al (2016) Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. Genome Biol 17:191. https://doi.org/10.1186/s13059-016-1053-6

Starnawska A, Tan Q, McGue M, Mors O, Børglum AD, Christensen K, Nyegaard M, Christiansen L (2017) Epigenome-wide association study of cognitive functioning in middle-aged monozygotic twins. Front Aging Neurosci 9:413. https://doi.org/10.3389/fnagi.2017.00413

Straussman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N, Simon I, Yakhini Z, Cedar H (2009) Developmental programming of CpG island methylation profiles in the human genome. Nat Struct Mol Biol 16:564–571. https://doi.org/10.1038/nsmb.1594

Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 8:25. https://doi.org/10.1186/1471-2105-8-25

Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. BMC Bioinformatics 9:307. https://doi.org/10.1186/1471-2105-9-307

Strobl C, Malley J, Tutz G (2009) An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol Methods 14:323–348. https://doi.org/10.1037/a0016973

Suderman M, McGowan PO, Sasaki A, Huang TCT, Hallett MT, Meaney MJ, Turecki G, Szyf M (2012) Conserved epigenetic sensitivity to early life experience in the rat and human hippocampus. Proc Natl Acad Sci U S A 109(Suppl 2):17266–17272. https://doi.org/10.1073/pnas.1121260109

Svane AM, Soerensen M, Lund J, Tan Q, Jylhävä J, Wang Y, Pedersen N, Hägg S, Debrabant B, Deary I, Christensen K, Christiansen L, Hjelmborg J (2018) DNA methylation and all-cause mortality in middle-aged and elderly Danish twins Genes (Basel) 9 https://doi.org/10.3390/genes9020078

Sziráki A, Tyshkovskiy A, Gladyshev VN (2018) Global remodeling of the mouse DNA methylome during aging and in response to calorie restriction. Aging Cell 17:e12738. https://doi.org/10.1111/acel.12738

Szyf M, Bick J (2013) DNA methylation: a mechanism for embedding early life experiences in the genome. Child Dev 84:49–57. https://doi.org/10.1111/j.1467-8624.2012.01793.x

Talens RP, Christensen K, Putter H, Willemsen G, Christiansen L, Kremer D, Suchiman HED, Slagboom PE, Boomsma DI, Heijmans BT (2012) Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. Aging Cell 11:694–703. https://doi.org/10.1111/j.1474-9726.2012.00835.x

Toyoda S, Kawaguchi M, Kobayashi T, Tarusawa E, Toyama T, Okano M, Oda M, Nakauchi H, Yoshimura Y, Sanbo M, Hirabayashi M, Hirayama T, Hirabayashi T, Yagi T (2014) Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity. Neuron 82:94–108. https://doi.org/10.1016/j.neuron.2014.02.005

Triche TJ Jr, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD (2013) Low-level processing of Illumina Infinium DNA methylation beadarrays. Nucleic Acids Res 41:e90. https://doi.org/10.1093/nar/gkt090

Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, Thurman RE, Kaul R, Myers RM, Stamatoyannopoulos JA (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res 22:1680–1688. https://doi.org/10.1101/gr.136101.111

Weidner CI, Wagner W (2014) The epigenetic tracks of aging. Biol Chem 395:1307–1314. https://doi.org/10.1515/hsz-2014-0180

Weiner JA, Jontes JD (2013) Protocadherins, not prototypical: a complex tale of their interactions, expression, and functions. Front Mol Neurosci 6:4. https://doi.org/10.3389/fnmol.2013.00004

Willemsen G, Ward KJ, Bell CG, Christensen K, Bowden J, Dalgård C, Harris JR, Kaprio J, Lyle R, Magnusson PKE, Mather KA, Ordoňana JR, Perez-Riquelme F, Pedersen NL, Pietiläinen KH, Sachdev PS, Boomsma DI, Spector T (2015) The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: the Discordant Twin (DISCOTWIN) Consortium. Twin Res Hum Genet 18:762–771. https://doi.org/10.1017/thg.2015.83

Wilson AS, Power BE, Molloy PL (2007) DNA hypomethylation and human diseases. Biochim Biophys Acta 1775:138–162. https://doi.org/10.1016/j.bbcan.2006.08.007

Wong CCY, Caspi A, Williams B, Craig IW, Houts R, Ambler A, Moffitt TE, Mill J (2014) A longitudinal study of epigenetic variation in twins. Epigenetics 5:516–526. https://doi.org/10.4161/epi.5.6.12226

Zhang X, Takata K, Cui W, Miyata-Takata T, Sato Y, Noujima-Harada M, Yoshino T (2016) Protocadherin gamma A3 is expressed in follicular lymphoma irrespective of BCL2 status and is associated with tumor cell growth. Mol Med Rep 14:4622–4628. https://doi.org/10.3892/mmr.2016.5808