

# Properties of the epigenetic clock and age acceleration

Louis El Khoury<sup>1\*\*</sup>, Tyler Gorrie-Stone<sup>1</sup>, Melissa Smart<sup>2</sup>, Amanda Hughes<sup>2</sup>, Yanchun Bao<sup>2</sup>, Alexandria Andrayas<sup>1</sup>, Joe Burrage<sup>3</sup>, Eilis Hannon<sup>3</sup>, Meena Kumari<sup>2</sup>, Jonathan Mill<sup>3</sup>, Leonard C Schalkwyk<sup>1\*</sup>

1) School of Biological Sciences, University of Essex, Colchester UK

2) Institute for Social and Economic Research, University of Essex, Colchester UK

3) Medical School, University of Exeter, Exeter UK

\* Corresponding

\*\* Current Address: Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester MN USA

LEK elkhoury.louis@mayo.edu

TGS tgorri@essex.ac.uk

MS mcsmar@essex.ac.uk

AH a.hughes@essex.ac.uk

YB ybaoa@essex.ac.uk

AA aandra@essex.ac.uk

JB J.Burrage@exeter.ac.uk

EH E.J.Hannon@exeter.ac.uk

MK mkumari@essex.ac.uk

JM J.Mill@exeter.ac.uk

LS lschal@essex.ac.uk

## Abstract

**Background:** The methylation status of numerous CpG sites in the human genome varies with age. The Horvath epigenetic clock used a wide variety of published DNA methylation data to produce an age prediction that has been widely used to, predict age in unknown samples, and draw conclusions about speed of ageing in various tissues, environments, and diseases. Despite its utility, there are a number of assumptions in the model that require examination. We explore the characteristics of the model in whole blood and multiple brain regions from older people, who are not well represented in the original training data, and in blood from a cross-sectional population study.

**Results:** We find that the model systematically underestimates age in tissues from older people. A decrease in slope of the predicted ages were observed at approximately 60 years, indicating that some loci in the model may change differently with age, and that age acceleration measures will themselves be age-dependent. This is seen most strongly in the cerebellum but is also present in other examined tissues, and is consistently observed in multiple datasets. An apparent association of Alzheimer's disease with age acceleration disappears when age is used as a covariate. Association tests in the literature use a variety of methods for calculating age acceleration and often do not use age as a covariate. This is a potential cause of misleading findings.

**Conclusions:** Associations of phenotypes with age acceleration should be evaluated cautiously, and chronological age should be included as a covariate in all analyses.

**Keywords:** DNA Methylation, epigenetic clock, age acceleration

## Background

Cellular differentiation and growth are orchestrated by epigenetic modifications. DNA methylation is the most stable and easily assayed epigenetic mark and it is indicative of many changes that happen throughout life. Different cell types have dramatically different DNA methylation profiles and there is a substantial literature on CpG loci whose DNA methylation status is associated with participant age in cross sectional studies (Hannum *et al.*, 2013; Horvath, 2013; Spiers *et al.*, 2015; Zbieć-Piekarska *et al.*, 2015). These sites thus appear to change over time, which, amongst other processes, could reflect developmental changes, cumulative environmental effects, and changes in cell-type composition. Identifying these sources of variation could give insights into multiple age-related processes, and also provide a way of estimating the age of study participants at the time of sample collection. There are changes to the epigenome with age, the majority of studies have shown global demethylation (Wilson *et al.*, 2007; Bjornsson *et al.*, 2008; Zampieri *et al.*, 2015), although there is hypermethylation in CpG islands within gene promoters (Bell *et al.*, 2012).

Horvath (2013) used a large collection ( $n > 8000$ ) of publically available Illumina HumanMethylation array data on multiple tissue types to train and test a model for age prediction from 353 CpG loci. This 'epigenetic clock' is extremely valuable as a way of estimating study participant ages, and possibly as an indicator of whether there are alterations in the ageing rate of certain tissues or individuals. With this in mind, it is attractive to try to relate these epigenetic changes with age related conditions such as dementia and Alzheimer disease (AD) (Zampieri *et al.*, 2015), frailty (Breitling *et al.*, 2016) and all-cause mortality (Marioni *et al.*, 2015). However, as the model was derived using data collected at a single time point and does not include within person changes it therefore does not directly assess ageing.

Ageing is a complex process described as a slow and gradual decline in physiological integrity leading to diminished function (López-Otín *et al.*, 2013). Many factors such as genomic instability, epigenetic alterations, mitochondrial dysfunctions and others, have been described by Lopez-Otin *et al.* (2013) as the hallmarks of ageing. Although these changes are involved with ageing and longevity, it is still unclear how these modifications are affected by the cells' and tissues' interaction

with the environment. Although the epigenetic clock developed by Horvath (2013) provides an estimate of age, the testing data used in generating this clock did not have a large representation of brain tissue from elderly individuals and as such it is unclear if the clock is accurate in older age groups, or those with age-related diseases.

We have previously published an epigenome-wide association study (EWAS) in AD, utilizing four brain tissues and pre-mortem blood, and demonstrated DNA methylation differences at specific loci in a tissue-specific manner (Lunnon *et al.*, 2014). This dataset offers an ideal opportunity to investigate the properties of the Horvath (2013) model on different tissues in both elderly non-demented individuals and AD sufferers. We further explore the properties of the model using a cross sectional population sample from the *Understanding Society* study, which has a wide range of ages.

Because the AD effects found in Lunnon *et al* (2014) and De Jager *et al* (2014) are quite small in magnitude and in a set of probes not overlapping with the ones used by the Horvath (2013) model, we did not expect to see 'age acceleration' associated with AD. For there to be an association the loci included in the model would be exhibiting a subtle but pervasive AD effect, too small to be identifiable on a probewise analysis, but perhaps reflective of a common variable-rate biological ageing clock mechanism. Age acceleration associated with disease states or environmental factors might support the notion such a mechanism does exist, but we would like to caution that in a field where many variables are being tested against age acceleration (which can also be defined in several ways) one would expect a number of positive results by chance.

## Results

We first used the Horvath Epigenetic Clock to predict the DNAm ages of participants from two case-control studies of Alzheimer Disease (AD) included in our original publication, 390 brain and blood samples from 92 individuals in the London cohort (cohort 1) (Lunnon *et al.*, 2014) and 280 brain samples from 147 individuals in the Mount Sinai cohort (cohort 2) (Smith *et al.*, 2018). As shown in table 1, in these elderly individuals DNAm ages are consistently underestimated. As can be seen from figure 1, DNAm ages clearly correlated with actual ages, but strikingly for all tissues of the London cohort the slope is less than 1 (figure 1 A-F, dotted line).

This picture is further confirmed by the data from the UKHLS population study. Here blood samples are obtained from participants across the adult age span and confirm the observation that DNAm ages are increasingly underestimated with advanced age (figure 1G). In figure 2 we show that the same effect holds across a variety of brain tissue data sets listed in table 3. In particular, the cerebellum is severely under predicted; far more than other brain and blood tissues (figure 2). Our consistent findings in tissue samples, from patients and a population study of healthy participants indicates that the results are robust and not specific to population or sample characteristics.

The Horvath (2013) model includes 353 coefficients, but most of the magnitudes are clustered near 0 (figure 3a), leading us to hypothesize that the few largest coefficients might be particularly influential. In figure 3b, the distribution of DNA methylation proportions in cerebellum of the London cohort data is shown in a stripchart for each of the 157 loci with positive coefficients in the Horvath (2013) model, sorted by magnitude, with the largest coefficients at the right, and figure 3c for the negative coefficients (smallest at the left). It is striking that several loci with the largest positive and negative age coefficients in the model have low variance in this sample and also low DNA methylation.

One possible mechanism for the attenuated association of many CG sites in the model with age would be saturation (i.e. individual loci with positive coefficients reaching full methylation or negative coefficient ones reaching zero). To investigate this, we estimate the influence of each locus on the age estimate by dividing the coefficient from Horvath (2013) by an index of dispersion from our data, the

coefficient of variation. In figure 3d we explore this value, and see that the most influential probes (black points at the top and bottom of the plot) include examples of both small variance (and large coefficient) and large variance (and small coefficient). This is further explored in figure 3e in which the ten loci with the highest influence are shown, with the ages equivalent to the length of each line protruding out of each circle and forming the rug on the right hand side of each strip chart. The rugs, scaled to best reflect the age range in the space available, give an indication of the noisiness of the relationship of the individual betas with age: the age relationship is not immediately visible for individual probes. The coefficients are shown in the centre, and the sign is also denoted by the direction of the stripchart, downward facing triangles for negative. This shows that of these ten probes there are three (cg12830694, cg24580001, and cg02580606) that might be candidates for saturation because they are highly methylated and expected to increase further with age. We thus fitted a regression line between chronological age and the beta values of each of the 353 loci and plotted the slopes against the Horvath coefficients. The 157 positively correlated and the 196 negatively correlated CpGs with age in Horvath (2013) are plotted and the ten most influential loci are highlighted in black (fig 3f). Of the ten most influential loci, four have a slope opposite in sign to the Horvath coefficient where three loci (cg08090772, cg03019000, and cg04268405) with negative Horvath coefficients have a positive slope in the cerebellum, and one locus (cg24580001) with a positive Horvath coefficient displays a negative slope in the cerebellum. In the main, the larger coefficients demonstrate the same direction of effect in our data as in the Horvath model. Admittedly this analysis is based on one tissue and age range, but it appears that many of the smaller coefficients may be effectively random with no biological meaning, and present in the model due to overfitting.

It is worth noting that the cerebellum is characterised by elevated levels of 5-hydroxymethylcytosine (5hmC) (Lunnon et al., 2016). We found that 31 out of the 353 Horvath clock sites were amongst the 65663 elevated 5hmC probes found in the cerebellum by Lunnon *et al* (2016). Of these, two sites (cg04268405, and cg24580001) are among the most influential sites reported in figure3e. Given that 5hmC is not distinguished from 5mC following bisulfite conversion, it is possible that

age-associated changes to the 31 5hmC sites in the Horvath algorithm are offsetting the age predictions.

Finally, we examined whether age acceleration (calculated as the difference between DNAm age and chronological age) associates with AD neuropathology (measured using Braak score). Our results do show a weak association in some brain tissues (table 2). However, when age is included as a covariate, the association between age acceleration and AD pathology disappears (table 2). We also see this in the Mount Sinai cohort where no correlation was found between age acceleration and amyloid plaque levels in either the PFC or the STG when age is included as a covariate (table 2).

## Discussion

The Horvath (2013) 'clock' underestimates age in tissues and blood from older individuals. Following several publications listing loci at which DNA methylation associates with age (Hernandez *et al.*, 2011; Koch and Wagner, 2011), Horvath (2013) used penalized regression to train a multi-locus, cross tissue predictor of age using a large collection of public data. This involved a nonlinear transformation of the data to prevent the huge changes of childhood and adolescence from dominating the model. The resulting 'epigenetic clock' has been of practical use in predicting the age of unknown samples and as a quality check in epigenetic research. Additional widely used age predictors specific for blood were published by Hannum (2013) and Levine (2018) (phenotype-based). Here we analyse the Horvath (2013) clock, but the methods and many of the conclusions may be more widely applicable.

If the epigenetic clock is an index of an underlying ageing program that adapts to health and environment, then it resembles the circadian clock, and a biochemical mechanism for it remains to be discovered, particularly for adulthood and old age. Alternatively, it may represent the accumulation of DNA methylation change over time, analogous to a phylogenetic clock. When comparing DNA methylation profiles across tissues, individuals and other variables such as health, the dominant source of variation is the tissue, or more precisely the cell type. It is reasonable to suppose that this developmental blueprint can change over time in response to the environment, or simply drift or decay. This point of view corresponds roughly with the 'Epigenetic Maintenance' model posited by Horvath (2013).

Although the Horvath (2013) paper proposed an 'epigenetic maintenance' model of the epigenetic clock, which does not require an internal clock mechanism, the same work also featured the somewhat contradictory idea of 'age acceleration' in which discrepancies between DNA methylation (DNAm) age and chronological age might tell us something about a clock mechanism and the biological ageing status of the organism. These ideas are developed further by Horvath and Raj (2018). It is worth highlighting that in the latter third of the human age range, where you might most expect to see such associations, negative age acceleration increases with age. This means that any phenotype associated with age will appear to be associated with age

acceleration as well, and a correct analysis should include chronological age as a covariate.

In a broad but non- comprehensive survey of the literature (table S1), we observe a variety of methods of calculating age acceleration, and many studies that do not correct for chronological age. Initially  $\Delta$ -age (the difference between chronological age and the DNAm predicted age) was reported but alternative methods have since arisen: (1) the residual of regressing DNAm predicted age on chronological age (possibly in a model including covariates), (2) AgeAccel (difference between DNAm age value and the value predicted by a regression model in the control group), and (3) intrinsic (IEAA) and (4) extrinsic epigenetic age acceleration (EEAA) methods. Both IEAA and EEAA are methods applicable only on blood tissue since they subtract out the effect of blood cell count. One difference between the two, however, lies in that IEAA is based on the Horvath clock and does not take into consideration age-related decline of immune cells (Levine *et al.*, 2015). Contrarily, EEAA is defined on the basis of a weighted average of the epigenetic age from the Hannum clock and immune cells known to change with age (Horvath and Ritz, 2015).

In light of this methodological variety, we are concerned that the different epigenetic clocks, and the variety of age acceleration methods to choose from, would be laying a multiple testing trap that is overlooked by some colleagues in the field.

## Conclusion

Every adult experiences change over time in a way that makes the concept of 'biological age' compelling. The clock model has extremely interesting and useful characteristics but it is an extremely narrow summary of the DNA methylation profile. Many studies use age acceleration as the single epigenetic variable, based on only 353 CpG sites representing  $1.15 \times 10^{-5}\%$  of the methylome. Caution is urged when using the epigenetic clock, in particular in case control studies where studies should be carefully matched for age and may require further adjustment for age. Highly-powered epigenetic studies will continue to show some intriguing associations between age acceleration and varied health outcomes no matter which vision is true.

Finally, an ideal epigenetic clock and age predictor might be within reach but it would require the incorporation of all epigenetic changes observed with age and not just

DNA methylation. Until then, it would be recommendable and more accurate to refer to the existing clocks as DNA methylation clocks instead.

## Methods

### Samples

#### Tissue Samples:

Brain tissue samples (London cohort) were obtained from individuals diagnosed with Alzheimer's disease (AD, n = 61) and from non-demented elderly control individuals (CON, n = 31) through the MRC London Neurodegenerative Disease Brain Bank as described in Lunnon *et al* (2014). In total four brain regions were analyzed (prefrontal cortex (PFC), the enthorhinal cortex (EC), the superior temporal gyrus (STG), and the cerebellum (CER)) and pre-mortem blood from a subset of individuals, collected as part of the Biomarkers of AD Neurodegeneration study. A second independent cohort (Mount Sinai cohort) was obtained from the Mount Sinai Alzheimer's disease and Schizophrenia Brain Bank. This cohort consisted of two brain regions (PFC and STG) for 75 AD and 72 CON donors (Smith *et al.*, 2018).

#### Population sample: The UK Household Longitudinal study (UKHLS)

UKHLS is an annual household-based panel study which started collecting information about the social, economic and health status of its participants in 2009. Our analysis data set is drawn from one of the arms of UKHLS, namely, the British Household Panel Survey (BHPS), which merged with UKHLS in 2010 at the start of wave two. UKHLS collected additional health information, including blood samples for genetic and epigenetic analysis, at wave 3 for BHPS.

[\(\[www.understandingsociety.ac.uk\]\(http://www.understandingsociety.ac.uk\)\)](http://www.understandingsociety.ac.uk)

### Methylomic Profiling

DNA from the London cohort tissue samples were bisulfite-treated using Zymo EZ 96 DNA methylation kit (Zymo Research) according to the manufacturer protocol. DNA methylation levels were assessed on an Illumina HiScan System using the Illumina Infinium HumanMethylation450 BeadChip as previously described by Hannon *et al.*

(2015). Raw signal intensities and probes for the London cohort were extracted using Illumina Genome Studio software were transformed into beta values using the Bioconductor wateRmelon package (Pidsley *et al.*, 2013). These were later normalized using the method implemented in the Horvath (2013) script. Data is available from both cohorts under GEO accession numbers GSE59685 (London cohort) and GSE80970 (Mount Sinai cohort).

1193 DNA samples from UKHLS were bisulfite-treated using Zymo EZ 96 DNA methylation kit (Zymo Research) according to the manufacturer protocol. DNA methylation levels were assessed on an Illumina HiScan System (Illumina) using the Illumina Infinium Epic Methylation BeadChip and samples were randomly assigned to chips and plates to minimise batch effects. Furthermore, in order to resolve any experimental inconsistencies, and to approve data quality, a fully methylated control (CpG Methylated HeLa Genomic DNA; New England BioLabs, MA, USA) was included in a random position on each plate. Raw signal intensities and probes for UKHLS were extracted using Illumina Genome Studio software and were transformed into beta values using the Bioconductor bigmelon package ([10.18129/B9.bioc.bigmelon](#)). These were later normalized using dasen function from the wateRmelon package (Pidsley *et al.*, 2013). After QC a final n of 1175 was reached.

### DNA Methylation Age prediction

DNA methylation (DNAm) age was assessed for all samples of the London and Mt Sinai datasets on the the R statistical environment (R Development Core Team, 2015) using the script provided by Horvath (2013) as well as through the online DNAm Age Calculator (<https://dnamage.genetics.ucla.edu/>). These methods predicted age based on the DNAm coefficients of 353 CpG sites. The model (although not the custom normalization method) is also implemented in the agep() function of the wateRmelon package (version 1.17.0). This function was used to predict the ages of the UKHLS samples.

To assess the weighted average methylation levels, and to maximize the number of brain samples included in our assessment of age prediction, publically available

450KMethylation brain tissue datasets obtained from GEO (GSE40360, GSE53162, GSE59457, GSE61380, GSE61431, GSE67748, GSE67749, and GSE89702) along with the London and Mount Sinai cohorts were analysed (table 3).

## Declarations

-Ethics approval and consent to participate: London cohort- Ethical approval NHS was provided by South East London REC 3; Mount Sinai Cohort- Ethical approval for the project was provided by the University of Exeter Medical School Research Ethics Committee under application number 14/02/041; UKHLS- Participants gave informed written consent for their blood to be taken and stored for future scientific analysis. The UKHLS has been approved by the University of Essex Ethics Committee and the nurse data collection by the National Research Ethics Service (10/H0604/2).

-Availability of data and material: The data used in this publication is previously published and available under GEO accession numbers given in table 3. Individual DNA methylation levels are available on application through the European Genome-phenome Archive under accession EGAS00001001232 (<https://www.ebi.ac.uk/ega/home>). Specific details can be found here (<https://www.understandingsociety.ac.uk/about/health/data>). Phenotypes linked to DNA methylation data are available through application to the METADAC ([www.metadac.ac.uk](http://www.metadac.ac.uk))

-Competing interests: The authors declare no competing interests.

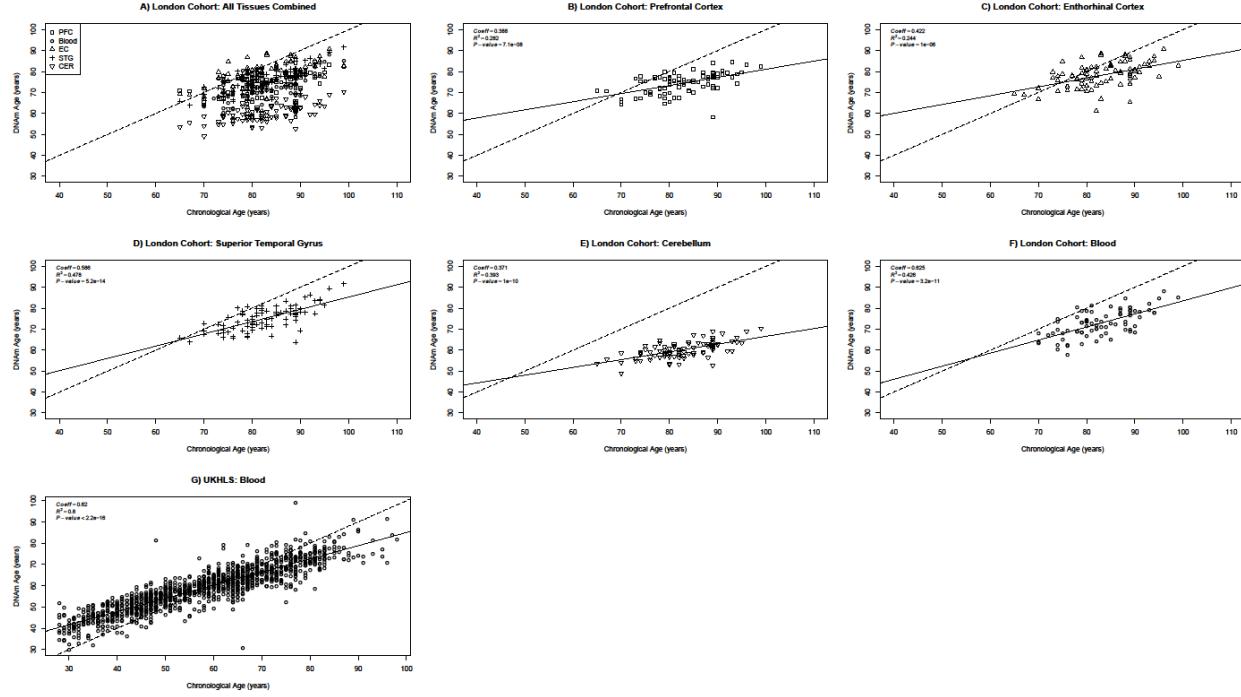
Funding: This work was supported by grants from the UK Medical Research Council (MRC) (grant numbers MR/K013807/1 and MR/L010674/1) to JM and LS. The AD data sets were produced under NIH grant R01 AG036039 to JM and LS. The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council (Grant Number: ES/M008592/1).

Authors' contributions: LS designed the study, LEK, TGS and LS did the analyses, MS AH YB AA JB EH and TGS generated and curated the UKHLS DNA methylation

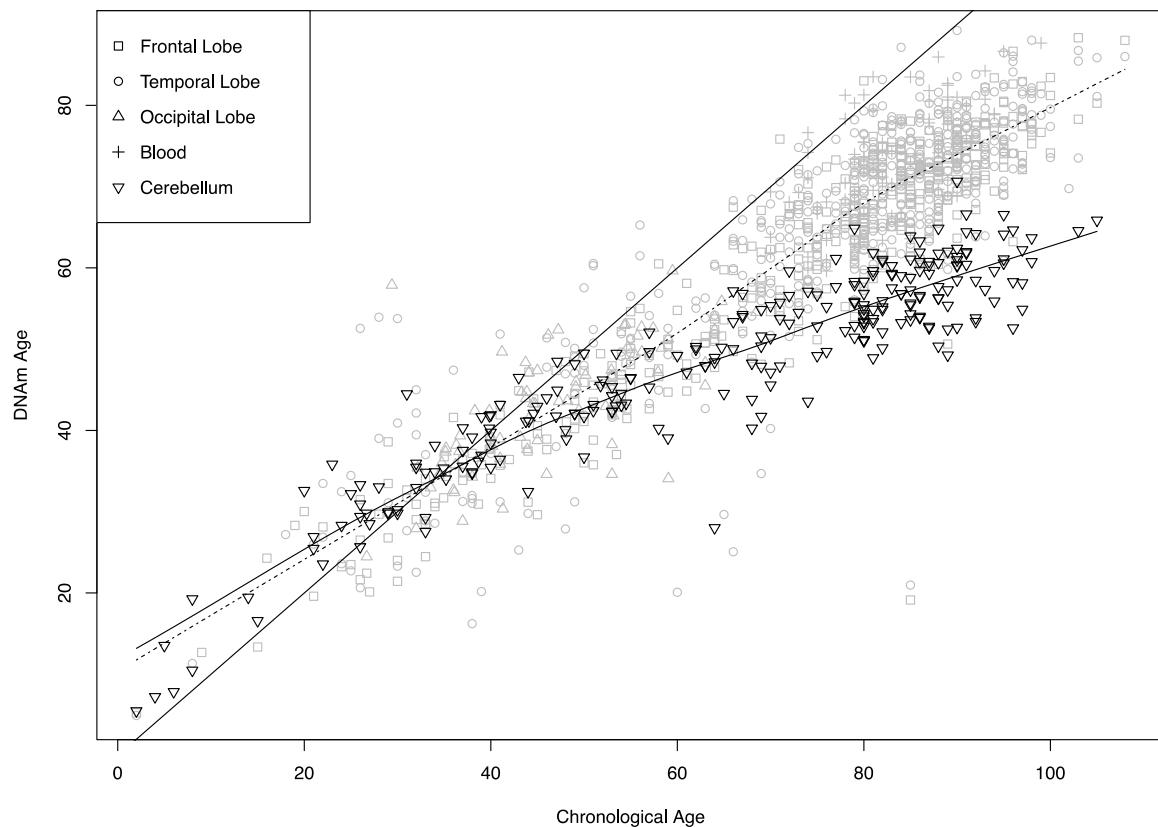
data, MK and JM advised and oversaw the work and all authors were involved in drafting and approving the manuscript.

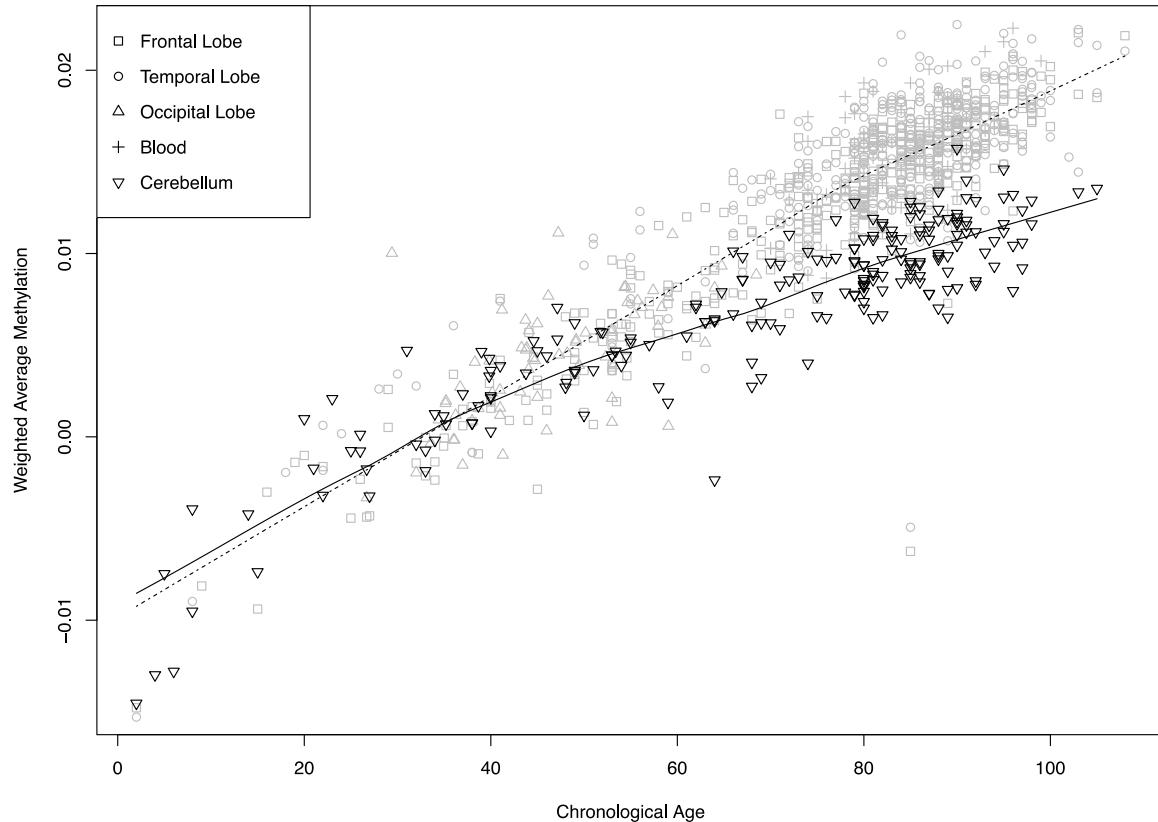
Acknowledgements: *Understanding Society*: Analysis was facilitated by access to the Genome high performance computing cluster at the University of Essex, School of Biological Sciences.

## Figures

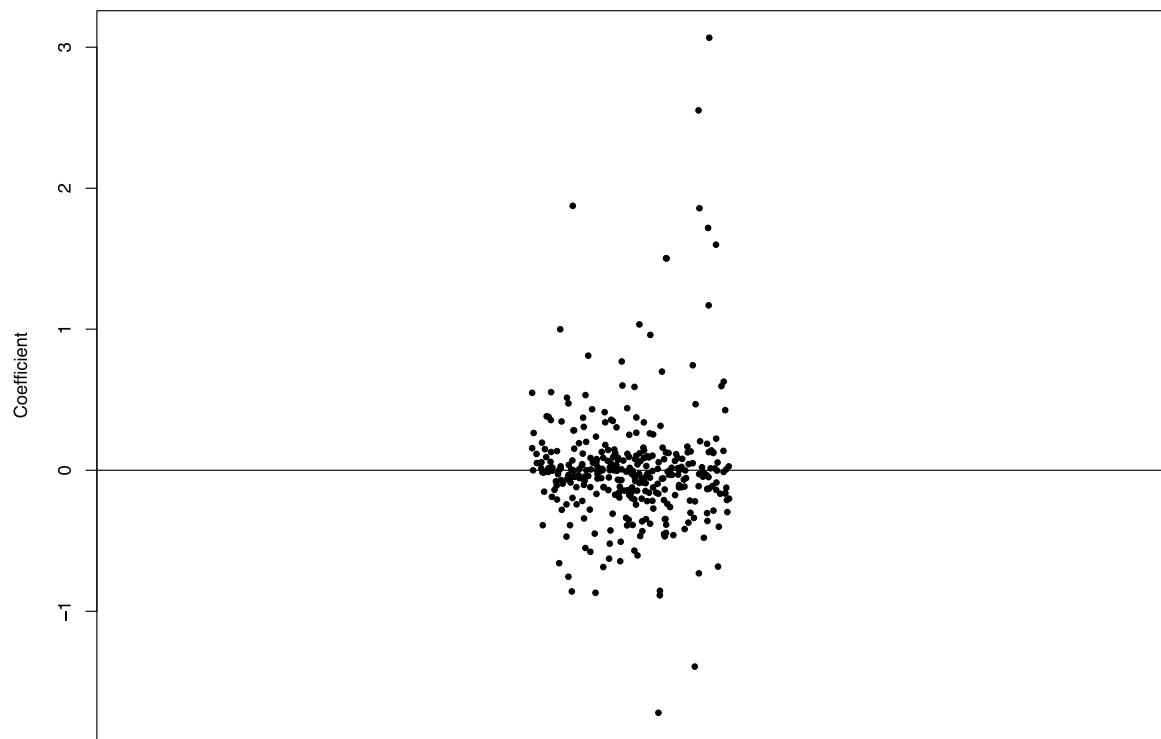


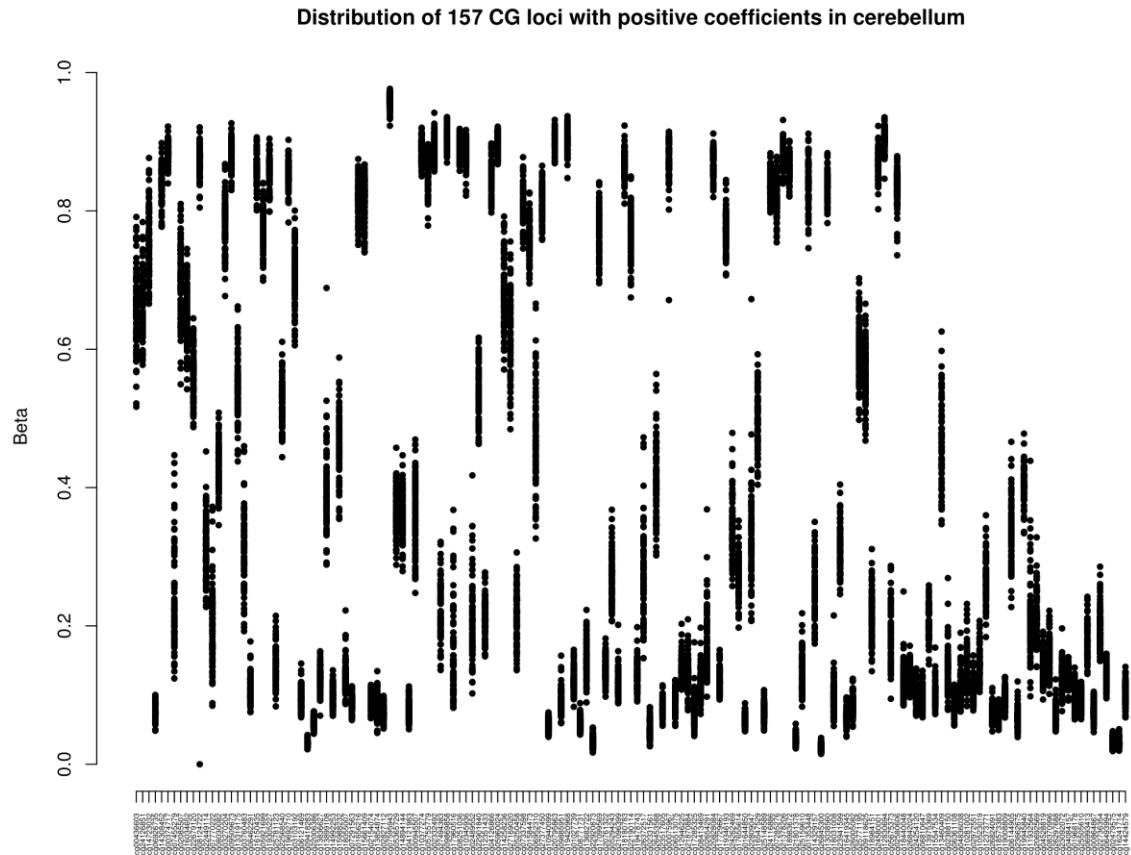
**Figure 1 Scatterplot of Chronological vs DNAm ages of brain and blood samples.** Each point corresponds to an independent sample. The dotted line is the  $y=x$  bisector line and the solid lines correspond to the regression line of each tissue. PFC: prefrontal cortex, EC: entorhinal cortex, STG: the superior temporal gyrus, CER: cerebellum (data from Lunnon *et al* 2014 for panels A-F and Gorrie-Stone *et al* (2018, under peer-review) for panel G).

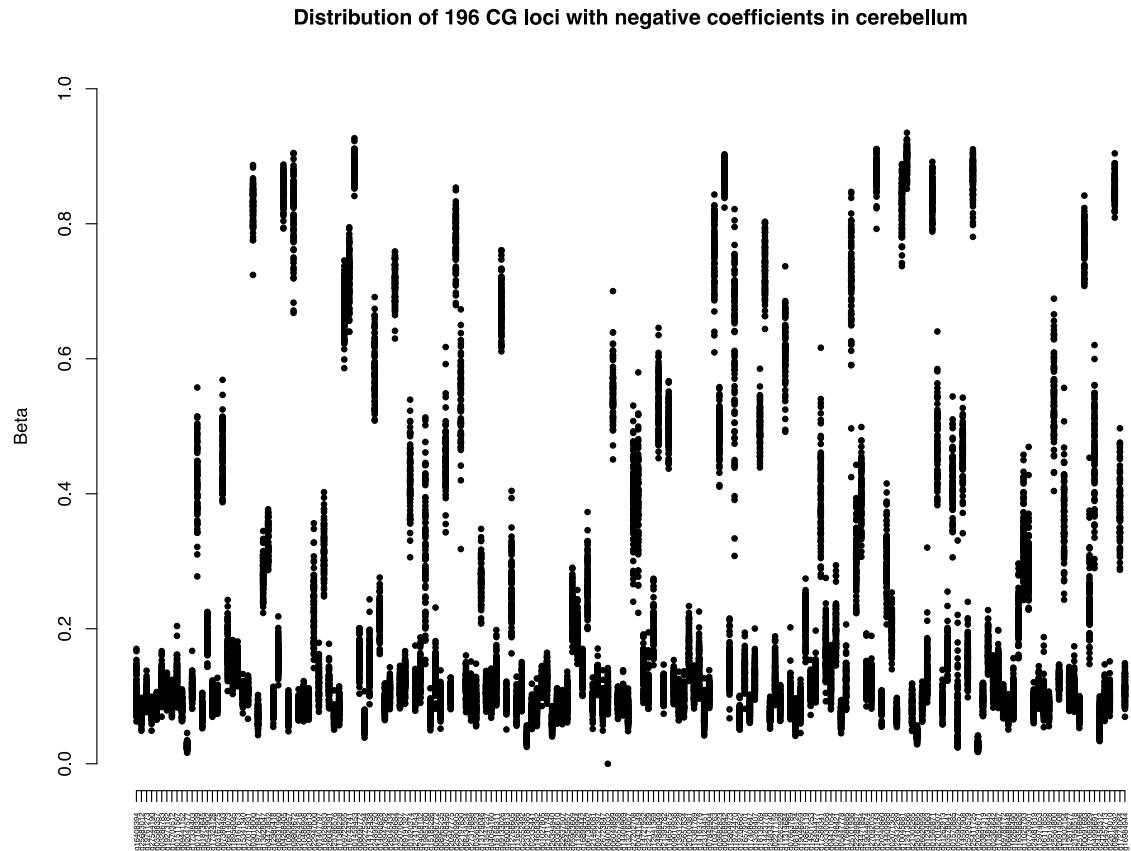


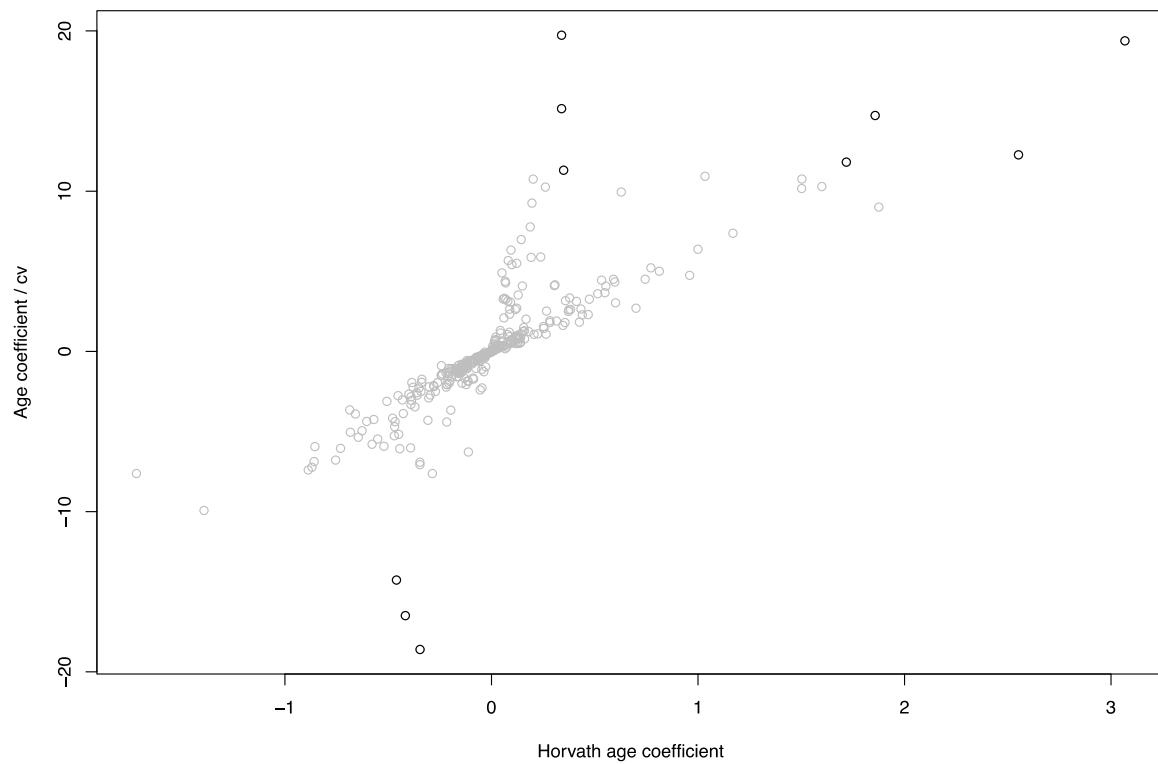


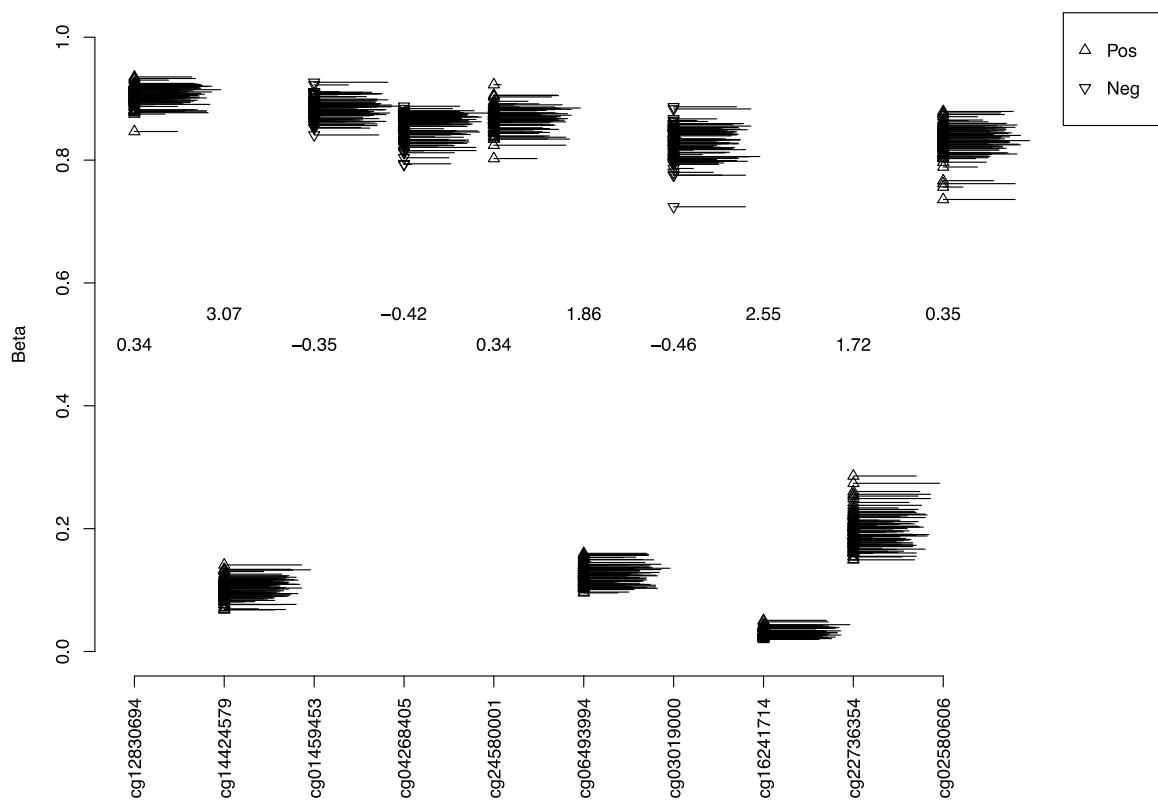
**Figure 2 A)** Exploration of Chronological vs DNAm ages and **B)** weighted average DNA methylation of the 353 Horvath clock CpGs vs chronological age in a dataset combining GSE40360, GSE53162, GSE59457, GSE61380, GSE61431, GSE67748, GSE67749, GSE89702, the London, and the Mount Sinai cohorts (Frontal lobe n=446, temporal lobe n=504, occipital lobe n=72, cerebellum n=246, blood n=80). An increase in divergence is observed as chronological age increases and a reduced slope among all samples (dotted curve) and especially the cerebellum samples (solid curve) can be noticed starting from the age of 60 which suggests the importance of introducing another inflection in the model at around that time. The solid line in fig 2a is the  $y=x$  bisector line.

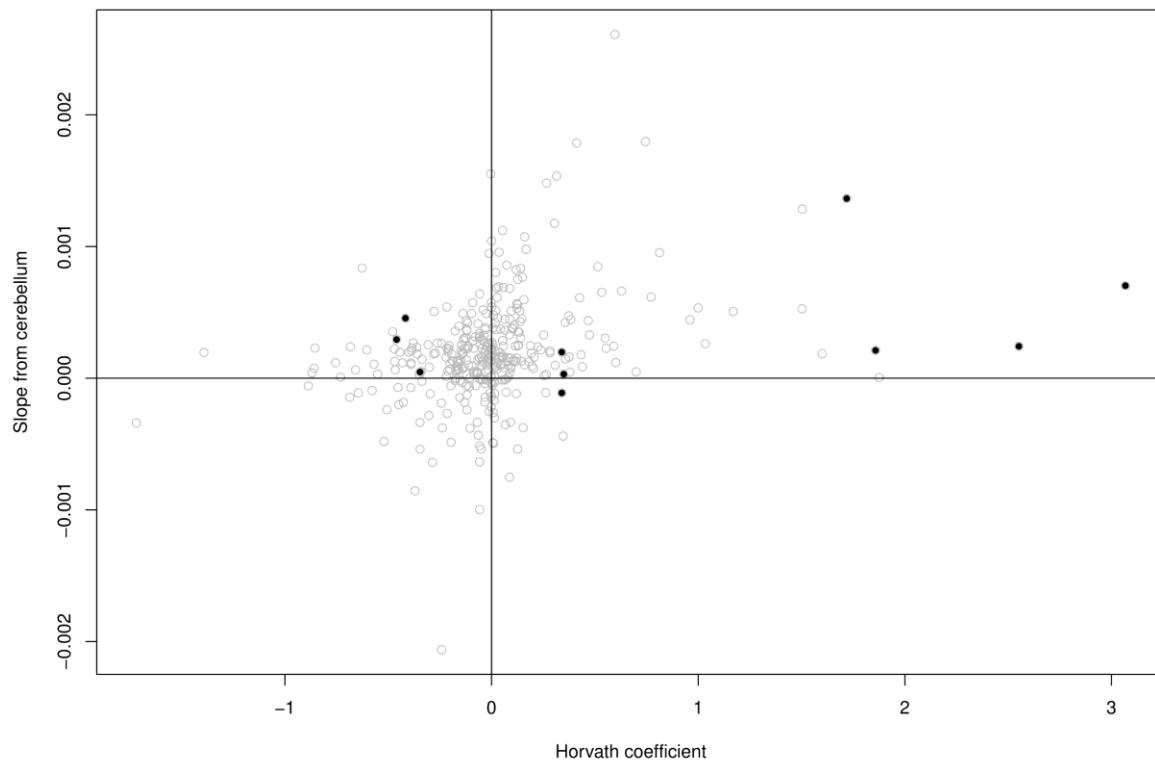












**Figure 3 Exploration of model coefficients in the elderly cerebellum** a) stripchart showing the distribution of the 353 coefficients from Horvath (2013), b) stripchart showing the distribution of DNA methylation values of the 158 CG loci with positive coefficients in 83 cerebellum samples, c) stripcharts showing the distribution of DNA methylation values of the 195 CG loci with negative coefficients in 83 cerebellum samples, d) Scatter plot of age coefficients against their influence score (coefficient from Horvath (2013)/coefficient of variation in our data). The ten most influential loci are plotted in black, e) ten most influential loci, with the ages represented as a rug on the right hand side of each strip chart. The Horvath coefficients are shown in the center, and their sign is also denoted by the direction of the triangles, upward facing for positive and downward facing for negative, f) scatter plot of Horvath (2013) coefficients against their linear-model age coefficients in our data. The 10 most influential probes are shown in black.

**Table 1** Chronological ages of samples obtained from the London Cohort

Tissue	London CON				London AD				Mt Sinai CON				Mt Sinai AD			
	N	CHRN Age	DNAm Age	N	CHRN Age	DNAm Age	N	CHRN Age	DNAm Age	N	CHRN Age	DNAm Age	N	CHRN Age	DNAm Age	
Prefrontal Cortex	24	78.62 ± 11.07	66.74 ± 7.57	60	86.40 ± 7.53	71.63 ± 5.49	68	82.76 ± 7.77	68.23 ± 9.26	74	88.53 ± 6.67	73.30 ± 5.39				
Entorhinal Cortex	21	79.62 ± 10.34	73.92 ± 8.58	58	86.50 ± 7.64	74.97 ± 6.49	-	-	-	-	-	-				
Superior Temporal Gyrus	26	76.73 ± 13.16	66.48 ± 9.76	61	86.34 ± 7.47	72.93 ± 6.08	70	83.21 ± 8.13	67.99 ± 8.59	74	88.40 ± 6.76	72.15 ± 6.61				
Cerebellum	23	76.00 ± 13.56	54.06 ± 7.23	60	86.15 ± 7.38	57.61 ± 4.75	-	-	-	-	-	-				
Blood	9	80.00 ± 6.14	70.74 ± 6.79	48	83.83 ± 7.00	74.99 ± 7.42	-	-	-	-	-	-				

Values are expressed as means ± SD. The total number of subjects (n) is in parentheses.

CHRN age: Chronological age

DNAm age: epigenetic age predicted according to the Horvath (2013) model.

**Table 2** Regression analysis of epigenetic age acceleration of four brain tissues and blood from the London Brain Bank cohort versus brain Braak stage and of two brain tissues from the Mount Sinai cohort versus amyloid plaque levels. Both analyses were corrected for the chronological ages.

	London Cohort		London Cohort (age in model)		Mount Sinai Cohort (age in model)	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Prefrontal Cortex	0.434	0.164	-0.345	0.107	-0.029	0.629
Enthorhinal Cortex	0.969	0.004	0.424	0.125	-	-
Sup. Temporal Gyrus	0.653	0.019	-0.062	0.786	0.004	0.944
Cerebellum	1.127	0.002	-0.059	0.788	-	-
Blood	-0.278	0.632	-0.090	0.870	-	-

Shown for each cohort the coefficient and p-value for the regression analysis between age acceleration and Braak stages (London cohort) or amyloid plaque levels (Mount Sinai cohort) with chronological age as a covariate.

**Table 3** Breakdown of the phenotypes, tissues, and ages of the additional brain datasets

GEO Accession	Phenotype	Tissue	N	Age Range	Age (mean $\pm$ SD)	Reference
GSE40360	Multiple Sclerosis	White matter	27 cases	35 – 78	55.33 $\pm$ 9.99	Huynh <i>et al.</i> , 2014
			19 control	53 – 81	66.32 $\pm$ 8.50	
GSE53162	Autism	Temporal cortex	6 cases	2 – 51	21.83 $\pm$ 17.42	Ladd-Acosta <i>et al.</i> , 2014
			10 controls	22 – 56	39.6 $\pm$ 13.06	
		Prefrontal cortex	6 cases	2 – 29	17.5 $\pm$ 10.82	
			5 controls	19 – 56	34.8 $\pm$ 15.02	
		Cerebellum	7 cases	4 – 39	14.00 $\pm$ 12.34	
			6 controls	2 – 26	16.67 $\pm$ 10.03	
GSE59457	HIV	Cerebellum	20 cases	26 – 64	43.79 $\pm$ 9.49	Horvath and Levine, 2015
			4 controls	38 – 63	51.75 $\pm$ 10.31	
		Frontal lobe	2 cases	44 – 48	46.35 $\pm$ 2.48	
			4 controls	38 – 63	51.75 $\pm$ 10.31	
		Hippocampus	4 controls	38 – 63	51.75 $\pm$ 10.31	
			18 cases	26 – 64	44.34 $\pm$ 10.57	
		Medial frontal cortex	2 controls	38 – 53	45.5 $\pm$ 10.61	
			59 cases	26 – 68	49.89 $\pm$ 9.23	
		Midbrain	13 controls	32 – 64	49.69 $\pm$ 11.57	
			4 controls	38 – 63	51.75 $\pm$ 10.31	
GSE67748	HIV	Cerebellum	8 cases	27 – 64	44.00 $\pm$ 13.56	Horvath and Levine, 2015
			12 controls	15 – 85	66.67 $\pm$ 22.73	
GSE67749	HIV	Frontal lobe	8 cases	27 – 64	44.00 $\pm$ 13.56	Horvath and Levine, 2015
			25 controls	15 – 95	72.72 $\pm$ 17.80	
GSE61380	Schizophrenia	Prefrontal cortex	18 cases	24 – 73	45.50 $\pm$ 16.61	Pidsley <i>et al.</i> , 2014
			15 controls	21 – 69	42.20 $\pm$ 14.85	
GSE61431	Schizophrenia	Cerebellum	21 cases	31 – 87	61.76 $\pm$ 16.61	Pidsley <i>et al.</i> , 2014
			23 controls	25 – 96	61.39 $\pm$ 17.15	
		Frontal Cortex	20 cases	32 – 87	62.05 $\pm$ 14.81	
			23 controls	25 – 96	62.04 $\pm$ 17.09	
GSE89702	Schizophrenia	Striatum	37 cases	24 – 87	55.11 $\pm$ 17.58	Viana <i>et al.</i> , 2016
			45 controls	21 – 96	56.67 $\pm$ 20.08	
		Hippocampus	14 cases	31 – 79	59.71 $\pm$ 13.59	
			13 controls	25 – 95	64.46 $\pm$ 18.90	
		Cerebellum	16 cases	24 – 70	45.69 $\pm$ 16.93	
			17 controls	21 – 72	44.59 $\pm$ 15.79	

## References

Bell, J.T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., Shin, S.-Y., Dempster, E.L., Murray, R.M., Grundberg, E., Hedman, A.K., Nica, A., Small, K.S., Dermitzakis, E.T., McCarthy, M.I., Mill, J., Spector, T.D., 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. doi:10.1371/journal.pgen.1002629

Bjornsson, H.T., Sigurdsson, M.I., Fallin, M.D., Irizarry, R.A., Aspelund, T., Cui, H., Yu, W., Rongione, M.A., Ekström, T.J., Harris, T.B., Launer, L.J., Eiriksdottir, G., Leppert, M.F., Sapienza, C., Gudnason, V., Feinberg, A.P., 2008. Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 299, 2877–83. doi:10.1001/jama.299.24.2877

Breitling, L.P., Saum, K.U., Perna, L., Schöttker, B., Holleczek, B., Brenner, H., 2016. Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clin. Epigenetics* 8, 1–8. doi:10.1186/s13148-016-0186-5

De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J., McCabe, C., Tang, A., Raj, T., Replogle, J., Brodeur, W., Gabriel, S., Chai, H.S., Younkin, C., Younkin, S.G., Zou, F., Szyf, M., Epstein, C.B., Schneider, J.A., Bernstein, B.E., Meissner, A., Ertekin-Taner, N., Chibnik, L.B., Kellis, M., Mill, J., Bennett, D.A., 2014. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci* 17, 1156–1163. doi:10.1038/nn.3786|[rhttp://www.nature.com/nn/journal/v17/n9/abs/nn.3786.html#supplementary-information](http://www.nature.com/nn/journal/v17/n9/abs/nn.3786.html#supplementary-information)

Hannon, E., Lunnon, K., Schalkwyk, L., Mill, J., 2015. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics*. doi:10.1080/15592294.2015.1100786

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.B., 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. doi:10.1016/j.molcel.2012.10.016

Hernandez, D.G., Nalls, M.A., Gibbs, J.R., Arepalli, S., van der brug, M., Chong, S., Moore, M., Longo, D.L., Cookson, M.R., Traynor, B.J., Singleton, A.B., 2011. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum. Mol. Genet.* 20, 1164–1172. doi:10.1093/hmg/ddq561

Horvath, S., 2013. DNA methylation age of human tissues and cell types. *Genome biol* 14, R115. doi:10.1186/gb-2013-14-10-r115

Horvath, S., Levine, A.J., 2015. HIV-1 Infection Accelerates Age According to the Epigenetic Clock. *J. Infect. Dis.* 212, 1563–73. doi:10.1093/infdis/jiv277

Horvath, S., Raj, K., 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19, 371–384. doi:10.1038/s41576-018-0004-3

Horvath, S., Ritz, B.R., 2015. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany. NY)*. 7, 1130–1142. doi:10.18632/aging.100859

Huynh, J.L., Garg, P., Thin, T.H., Yoo, S., Dutta, R., Trapp, B.D., Haroutunian, V., Zhu, J., Donovan, M.J., Sharp, A.J., Casaccia, P., 2014. Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. *Nat. Neurosci.* 17, 121–30. doi:10.1038/nn.3588

Koch, C.M., Wagner, W., 2011. Epigenetic-aging-signature to determine age in different tissues. *Aging (Albany. NY)*. 3, 1018–1027. doi:100395 [pii]

Ladd-Acosta, C., Hansen, K.D., Briem, E., Fallin, M.D., Kaufmann, W.E., Feinberg, A.P., 2014. Common DNA methylation alterations in multiple brain regions in autism. *Mol. Psychiatry* 19, 862–71. doi:10.1038/mp.2013.114

Levine, M.E., Hosgood, H.D., Chen, B., Absher, D., Assimes, T., Horvath, S., 2015. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany. NY)*. 7, 690–700. doi:10.18632/aging.100809

Levine, M.E., Lu, A.T., Quach, A., Chen, B.H., Assimes, T.L., Bandinelli, S., Hou, L., Baccarelli, A.A., Stewart, J.D., Li, Y., Whitsel, E.A., Wilson, J.G., Reiner, A.P., Aviv, A., Lohman, K., Liu, Y., Ferrucci, L., Horvath, S., 2018. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany. NY)*. 10, 573–591. doi:10.18632/aging.101414

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., Kroemer, G., 2013. The hallmarks of aging. *Cell* 153, 1194–217. doi:10.1016/j.cell.2013.05.039

Lunnon, K., Hannon, E., Smith, R.G., Dempster, E., Wong, C., Burrage, J., Troakes, C., Al-Sarraj, S., Kepa, A., Schalkwyk, L., Mill, J., 2016. Variation in 5-hydroxymethylcytosine across human cortex and cerebellum. *Genome Biol.* 17, 27. doi:10.1186/s13059-016-0871-x

Lunnon, K., Smith, R., Hannon, E., De Jager, P.L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R., Condliffe, D., Harries, L.W., Katsel, P., Haroutunian, V., Kaminsky, Z., Joachim, C., Powell, J., Lovestone, S., Bennett, D.A., Schalkwyk, L.C., Mill, J., 2014. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* 17, 1164–70. doi:10.1038/nn.3782

Marioni, R.E., Shah, S., McRae, A.F., Chen, B.H., Colicino, E., Harris, S.E., Gibson, J., Henders, A.K., Redmond, P., Cox, S.R., Pattie, A., Corley, J., Murphy, L., Martin, N.G., Montgomery, G.W., Feinberg, A.P., Fallin, M.D., Multhaup, M.L., Jaffe, A.E., Joehanes, R., Schwartz, J., Just, A.C., Lunetta, K.L., Murabito, J.M., Starr, J.M., Horvath, S., Baccarelli, A.A., Levy, D., Visscher, P.M., Wray, N.R., Deary, I.J., 2015. DNA methylation age of blood predicts all-cause mortality in

later life. *Genome Biol.* 16, 1–12. doi:10.1186/s13059-015-0584-6

Pidsley, R., Viana, J., Hannon, E., Spiers, H., Troakes, C., Al-Saraj, S., Mechawar, N., Turecki, G., Schalkwyk, L.C., Bray, N.J., Mill, J., 2014. Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. *Genome Biol.* 15, 483. doi:10.1186/s13059-014-0483-2

Pidsley, R., Y Wong, C.C., Volta, M., Lunnon, K., Mill, J., Schalkwyk, L.C., 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* 14, 293. doi:10.1186/1471-2164-14-293

Smith, R.G., Hannon, E., De Jager, P.L., Chibnik, L., Lott, S.J., Condliffe, D., Smith, A.R., Haroutunian, V., Troakes, C., Al-Sarraj, S., Bennett, D.A., Powell, J., Lovestone, S., Schalkwyk, L., Mill, J., Lunnon, K., 2018. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. *Alzheimer's Dement.* doi:10.1016/j.jalz.2018.01.017

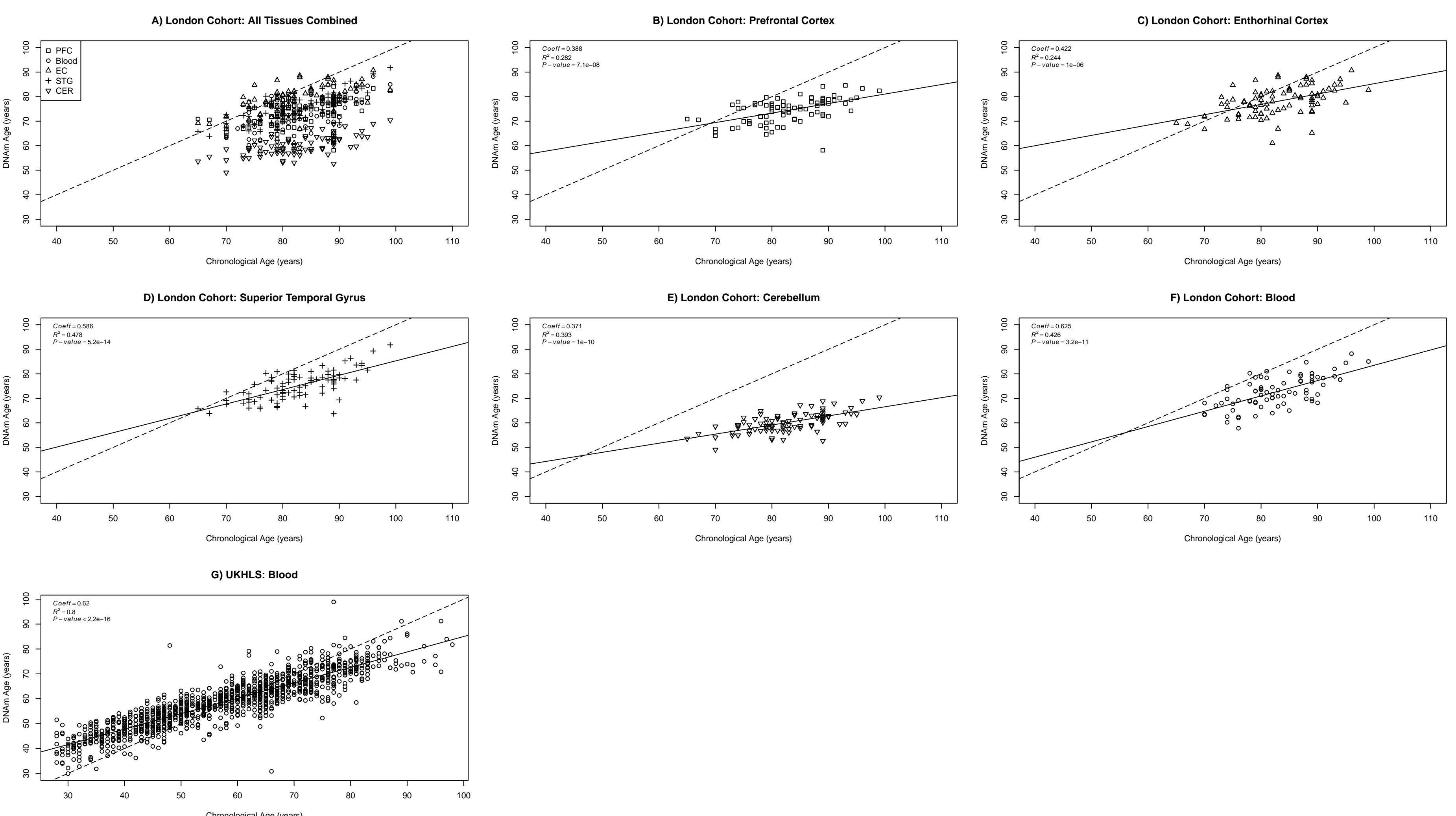
Spiers, H., Hannon, E., Schalkwyk, L.C., Smith, R., Wong, C.C.Y., O'Donovan, M.C., Bray, N.J., Mill, J., 2015. Methylomic trajectories across human fetal brain development. *Genome Res.* 25, 338–352. doi:10.1101/gr.180273.114

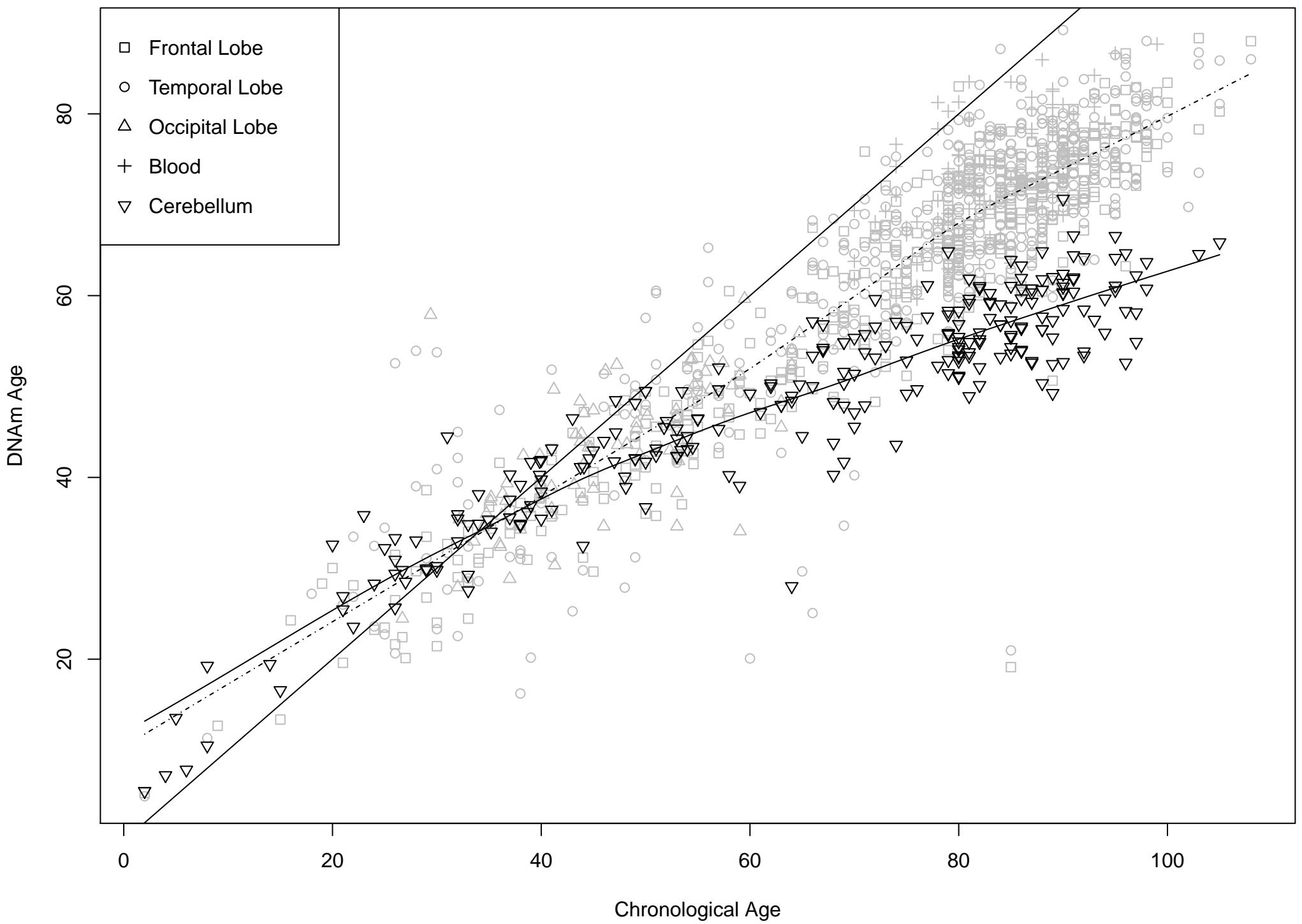
Viana, J., Hannon, E., Dempster, E., Pidsley, R., Macdonald, R., Knox, O., Spiers, H., Troakes, C., Al-Sarraj, S., Turecki, G., Schalkwyk, L.C., Mill, J., 2016. Schizophrenia-associated methylomic variation: molecular signatures of disease and polygenic risk burden across multiple brain regions. *Hum. Mol. Genet.* 62, ddw373. doi:10.1093/hmg/ddw373

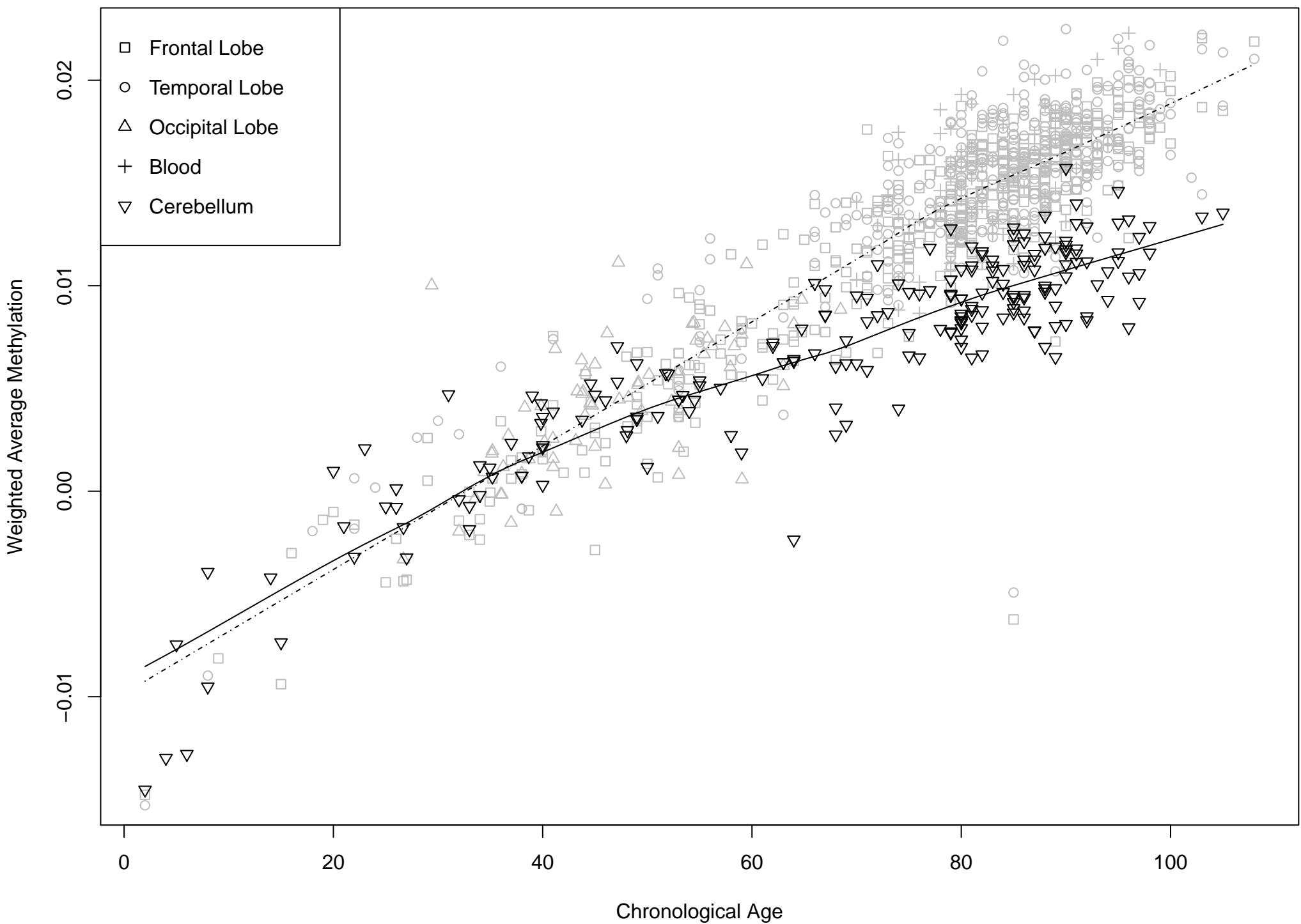
Wilson, A.S., Power, B.E., Molloy, P.L., 2007. DNA hypomethylation and human diseases. *Biochim. Biophys. Acta* 1775, 138–162. doi:10.1016/j.bbcan.2006.08.007

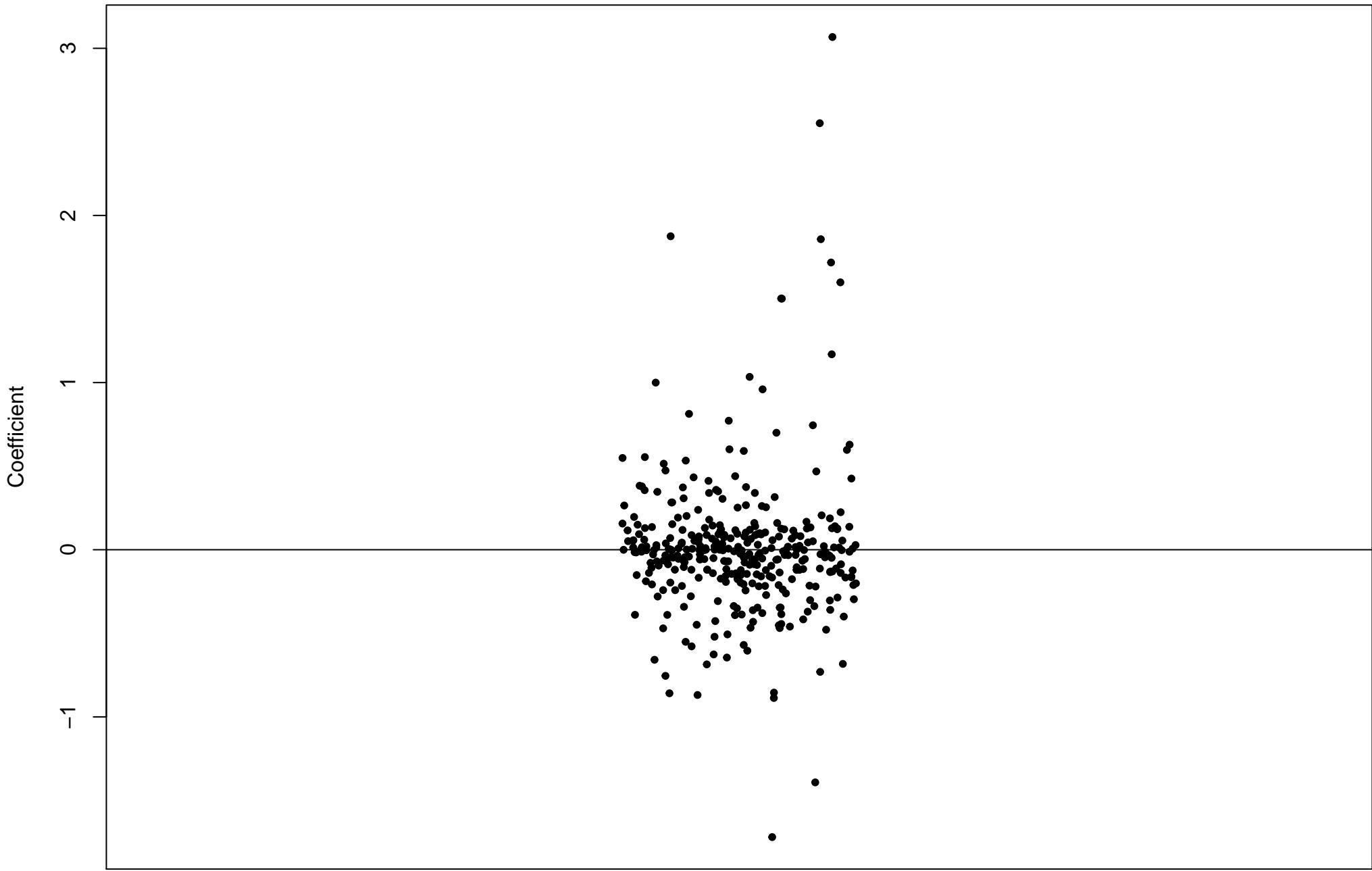
Zampieri, M., Ciccarone, F., Calabrese, R., Franceschi, C., Bürkle, A., Caiafa, P., 2015. Reconfiguration of DNA methylation in aging. *Mech. Ageing Dev.* 151, 1–11. doi:10.1016/j.mad.2015.02.002

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Makowska, Ż., Pałeczka, A., Kucharczyk, K., Płoski, R., Branicki, W., 2015. Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Sci. Int. Genet.* 17, 173–9. doi:10.1016/j.fsigen.2015.05.001

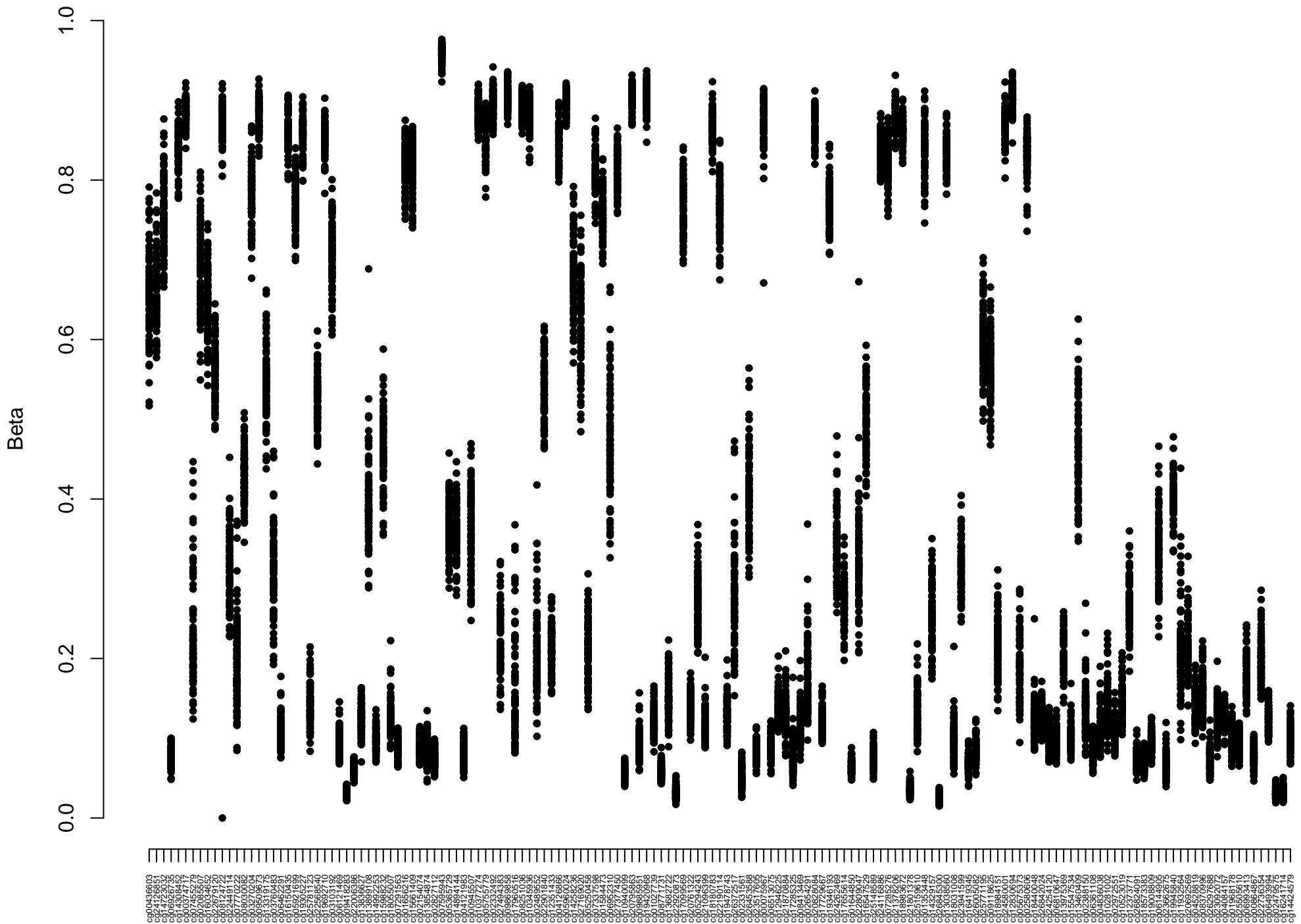








## Distribution of 157 CG loci with positive coefficients in cerebellum



# Distribution of 196 CG loci with negative coefficients in cerebellum

