Supplemental Material for

Genome-wide average DNA methylation is determined *in utero*

Shuai Li,[1] Ee Ming Wong,[2,3] Pierre-Antoine Dugué,[1,4] Allan F McRae,[5] Eunae Kim,[6] Ji-Hoon Eric Joo,[2,3] Tuong L Nguyen,[1] Jennifer Stone,[7] Gillian S Dite,[1] Nicola J Armstrong,[8] Karen A Mather,[9] Anbupalam Thalamuthu,[9] Margaret J Wright,[5] David Ames,[10] Roger L Milne,[1,4] Jeffrey M Craig,[11,12,13] Richard Saffery,[11,12] Grant W Montgomery,[14] Yun-Mi Song[15], Joohon Sung,[6,16] Timothy D Spector,[17] Perminder S Sachdev,[10] Graham G Giles,[1,4] Melissa C Southey,[2,3] John L Hopper[1,6,16*]

1. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville, Victoria, Australia

2. Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne, Parkville, Victoria, Australia

3. Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia

4. Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia

5. Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia

6. Department of Epidemiology, School of Public Health, Seoul National University, Seoul, South Korea

7. Centre for Genetic Origins of Health and Disease, Curtin University and the University of Western Australia, Perth, Western Australia, Australia

8. Mathematics and Statistics, Murdoch University, Perth, Western Australia, Australia

9. Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, Sydney, New South Wales, Australia

10. National Ageing Research Institute and University of Melbourne Academic Unit for Psychiatry of Old Age, Parkville, Victoria, Australia

11. Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia

12. Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia

13. School of Medicine, Deakin University, Geelong, Victoria, Australia

14. Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia

15. Department of Family Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

16. Institute of Health and Environment, Seoul National University, Seoul, South Korea

17. Department of Twin Research and Genetic Epidemiology, Kings College London, London, United Kingdom

* Corresponding author

Address: Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, 207 Bouverie Street, Parkville, Victoria 3010, Australia

Email: j.hopper@unimelb.edu.au

Phone: +61 3 8344 0697

METHODS

Studies and subjects

1. Peri/postnatal Epigenetic Twins Study (PETS)

The PETS is an Australian twin birth cohort aiming to study the plasticity of epigenetic marks during the intrauterine period and in early childhood.[1] A total of 250 newborn twin pairs were recruited between 2007 and 2009, and data on maternal factors during pregnancy, infant anthropometric measurements and biological specimens from different cell lineages were collected at several time points. Nine monozygotic (MZ) twin pairs and five dizygotic (DZ) twin pairs at birth, and six MZ pairs and four DZ pairs at age 18 months who had the Illumina Infinium HumanMethylation450 (HM450) BeadChip array DNA methylation data were included in this analysis. The study was approved by the Human Research Ethics Committees of the Royal Women's Hospital, Mercy Hospital for Women, and Monash Medical Centre, Melbourne. Written informed consent was obtained.

2. Brisbane Systems Genetic Study (BSGS)

The BSGS is an Australian study of twin families comprising adolescent twins, their siblings and their parents recruited into an ongoing study of the genetic and environmental factors influencing pigmented nevi and cognition.[2] 614 participants of European descent who had HM450 DNA methylation data were included in this analysis. The study was approved by the Human Research Ethics Committee of the Queensland Institute for Medical Research. Written informed consent was obtained.

3. Korean Healthy Twin Study (KHTS)

The KHTS is a study of twin families aiming to examine genetic and environmental factors underlying complex human diseases and traits in Korea.[3] Adult (30 years or older) same-sex twin pairs and their first-degree relatives were recruited. Questionnaire surveys were administered and biological specimens including blood were collected. 97 families were selected based on body mass index (BMI) or smoking status for DNA methylation research. 382 participants were included in this analysis. This study was

approved by the Institutional Review Board of Samsung Medical Centre and Busan Paik Hospital. All participants provided written informed consent.

4. Australian Mammographic Density Twins and Sisters Study (AMDTSS)

The AMDTSS is a twin family study conducted in Australia aiming to study mammographic density.[4] Between 2004 and 2009, female twin pairs aged 40–70 years who participated in the Australian Twins Study of Mammographic Density between 1995 and 1999 were asked to participate further, and their non-twin sisters were also invited to participate. Participants completed questionnaire surveys through telephone-administered interviews and donated blood samples. 479 participants were selected for DNA methylation research and were included in this analysis. The study was approved by the Human Research Ethics Committee of the University of Melbourne. All participants provided written informed consent.

5. Multiple Tissue Human Expression Resource (MuTHER) Study

The MuTHER is a study of middle-aged females, including 386 twin pairs and 84 singletons of European descent recruited through the TwinsUK Adult Twin Registry.[5] Punch biopsies (8mm) were taken from a photo-protected area adjacent and inferior to the umbilicus. Subcutaneous adipose tissue was dissected from each biopsy, weighted and immediately stored in liquid nitrogen. 246 twin pairs who had HM450 DNA methylation data were included in this analysis. The study was approved by the Research Ethics Committee of St. Thomas' Hospital, London. All participants provided written informed consent.

6. Older Australian Twins Study (OATS)

The OATS is a longitudinal, multi-centre study of twins aged 65 years and older that commenced in 2007 investigating healthy brain ageing.[6] Participants living in the three eastern states of Australia were administered a comprehensive face-to-face assessment including demographic, psychiatric, neuropsychological and medical measures. 108 MZ pairs who had HM450 DNA methylation data were included in this analysis. The study was approved by the ethics committees of the Australian Twin

Registry, University of New South Wales, University of Melbourne, Queensland Institute of Medical Research and the South Eastern Sydney and Illawarra Area Health Service. All participants provided written informed consent.

7. Melbourne Collaborative Cohort Study (MCCS)

The MCCS is a prospective cohort study of 41,514 healthy adult volunteers (24,469 women, 17,045 men) aged between 27 and 76 years (99.3% aged 40 − 69 years) recruited between 1990 and 1994.[7] Peripheral blood samples were obtained from participants at baseline. 5,629 participants from six nested cancer case-control studies were measured for DNA methylation. Among controls who had DNA methylation data, there were 43 spouse pairs. The spouse pairs were included in this analysis. The study was approved by the Cancer Council Victoria's Human Research Ethics Committee and performed in accordance with the institution's ethical guidelines. All participants provided written informed consent.

DNA methylation measurement

PETS

DNA was extracted from buccal cells. Buccal cells were collected with Catch-all Sample Collection Swabs (EPICENTRE Biotechnologies, Madison, WI, USA) and were stored at -20°C until DNA extraction. DNA was extracted with a standard phenol:chloroform method and bisulfite converted using the Methyl EasyXceed bisulfite modification kit (Human Genetic Signatures, North Ryde, Australia), according to the manufacturer's instructions. Twins from the same pair were processed in parallel. Raw intensity data were background corrected and methylation beta-values were generated using the R *minfi* package.[8] Data were pre-processed using the Illumina method within *minfi* and subset-quantile within-array normalization (SWAN)[9] was performed for combined normalization of Infinium type I and type II probes. Probes were removed if the average detection P value >0.001 in one or more samples, and/or on the sex chromosomes. Samples were excluded according to the average detection P value >0.05, poor bisulfite conversion efficiency and/or hierarchical clustering plots. See Martino *et al.*[10] for more details. Details for the methylation data in cord blood mononuclear cells measured by the HM27 assay

can be found in Gordon et al.[11] The datasets were obtained from Gene Expression Omnibus with the accession numbers GSE42700 and GSE36642.

BSGS

DNA was extracted from peripheral blood lymphocytes by the salt precipitation method.[12] Samples were randomly placed with respect to the chip they were measured on and to the position on that chip in order to avoid any confounding with family. Methylation scores for each CpG site are obtained as a ratio of the intensities of fluorescent signals and are represented as beta-values. Box-plots of the red and green intensity levels and their ratio were used to ensure that no chip position was under- or over-exposed, with any outlying samples repeated. Similarly, the proportion of probes with detection P value <0.01 was examined to confirm strong binding of the sample to the array. Probes exclusion criteria included on the sex chromosomes, having been annotated as binding to multiple chromosomes,[13] with zero CpG site and with more than 11 individuals with missing data or more than five individuals with detection P values >0.001. See McRae et al.[14] for more details. The dataset was obtained from Gene Expression Omnibus with the accession number GSE56105.

KHTS

DNA was extracted from peripheral blood lymphocytes. The measurement was conducted in two respective experiments (experiment I and II), with individuals from the same family included in the same experiment. For each experiment, quality control and data processing were performed separately, while the same analytic tools and methods were applied. The R package RnBeads[15] was applied to extract DNA methylation values across >485,000 CpG sites. For the quality control of the DNA methylation data, a series of probe and sample filtering steps were followed: probes mapping to sex chromosomes, associated with SNPs and/or out of CpG context were removed, and CpG probes and samples were filtered at detection P value of 0.01. The beta mixture quantile dilation (BMIQ) method[16] was used for normalization.

AMDTSS

DNA was extracted from dried blood spots stored on Guthrie cards using a method developed in-house.[17] DNA was sodium bisulfite converted using the EZ DNA Methylation-Gold protocol as per manufacturers' instructions (Zymo Research, Irvine, CA) and eluted in 20 µl elution buffer. DNA samples extracted from members of the same family were assayed on the same chip. Raw intensity data was processed by Bioconductor *minfi* package,[8] which included normalization of data using Illumina's reference factor-based normalization methods (*preprocessIllumina*) and the SWAN method[9] for type I and II probe bias correction. An empirical Bayes batch-effects removal method *ComBat*[18] was applied to minimise the technical variation across batches. Probes with detection *P* value >0.01 were assigned as missing. Probes with missing value in one or more samples were excluded, as were samples with >5% missing probes. See Li *et al.*[19] for more details.

MuTHER

DNA was extracted from adipose tissue samples. In order to avoid sampling biases, the included adipose tissue samples were randomized prior to DNA extraction. Genomic DNA was then isolated with a NORGEN DNA Purification Kit (Norgen Biotek Corporation) according to the manufacturer's protocol and quantified with PicoGreen. Raw data were imported to the GenomeStudio v.2010.3 software with the methylation module 1.8.2 for the extraction of the image intensities. Sample quality control based on probe detection and using the GenomeStudio *P* values of detection of signal above background. Probes that failed in at least one individual and that were not reported by the GenomeStudio software were discarded. The signal intensities for the methylated and unmethylated states were then quantile normalized for each probe type separately, and beta-values were calculated with R 2.12.0. See Grundberg *et al.*[5] for more details. The dataset was obtained from the ArrayExpress with the accession number E-MTAB-1866.

OATS

The data was generated at two respective experiments (experiment I and II). DNA was extracted from peripheral blood samples using either the Qiagen Autopure or a proteinase K method. Samples (co-twins) were randomised across the arrays. Raw intensity data were background corrected and

methylation beta-values were generated using the R *minfi* package.[8] The SWAN method[9] was performed for type I and II probe bias correction. Probes not detected in all samples were removed, as were probes containing SNPs and probes on the sex chromosomes.

MCCS

Samples in each case-control sub-study were processed separately during non-overlapping periods of time over a two-year period in the same laboratory with the same protocol. For the included 43 spouse pairs, DNA was extracted from peripheral blood samples collected at baseline, prior to any diagnosis of cancer, either from buffy coats (3%), lymphocytes (27%) or dried blood spots stored on Guthrie card (70%). Bisulfite conversion was performed using Zymo Gold single tube kit (EZ DNA Methylation-Gold kit, Zymo Research, CA, USA) according to the manufacturer's instructions. In order to minimize potential batch effects, matched cases and controls in each study were processed together and run on the same BeadChip and cancer subtypes were evenly distributed across the plates/chips. The same data pre-processing procedure was applied to each case-control sub-study, respectively. Raw intensity data was processed by Bioconductor *minfi* package,[8] which included normalization of data using Illumina's reference factor-based normalization methods (*preprocessIllumina*) and the SWAN method[9] for type I and II probe bias correction. Samples were excluded if >5% CpG probes (excluding chrX and chrY probes) had a detection *P* value >0.01, which were regarded as probes with 'missing value', while probes were excluded from further analysis if they had missing values for one or more samples. *ComBat*[18] was applied to the data from all samples across sub-studies to minimise the influence of chip effects. Beta-values after *ComBat* were used. See Severi *et al.*[20] and Wong *et al.*[21] for more details.

Statistical methods

Two-stage adjustment on GWAM

Within each study, we performed a two-stage adjustment on GWAM to minimize batch effects and to adjust for the effects of covariates. In the first stage, we applied the *ComBat* method[18] or a linear mixed effects model (Supplementary Table 1). *ComBat* was performed at the probe level during data

processing. The linear mixed effects model was fitted on GWAM with technical covariates as random effects. The residuals of the model were added to the mean of GWAM to give a 'batch-adjusted' GWAM. In the second stage, a linear regression model was used to adjust the 'batch-adjusted' GWAM for age, sex and study design or sampling factors (Supplementary Table 1).

Sensitivity analyses

Several sensitivity analyses were performed to examine the robustness of results to adjustment for cell mixture and to CpG selections. For study with methylation data from whole blood, proportions of monocytes, B cells, natural killers, CD4+ T cells, CD+8 T cells and granulocytes were estimated from the DNA methylation data using a reference-based method[22] by each study independently. For the PETS, the percentage of buccal epithelial cells was estimated using the "Buccal-Cell-Signature"[23]. For the MuTHER, a reference-free method[24] was used, and the dimension of the latent variable (i.e., number of cell types) was estimated to be one, using the function *EstDimRMT ()* from R package *isva* which applies random matrix theory[25]. Therefore, no cell mixture was adjusted for the MuTHER. In each study, familial correlations in GWAM additionally adjusted for associated cell type proportions were estimated. In the BSGS, the reference-based method[24] was also used to estimate cell mixture, and similar familial correlations were found (data not shown). GWAM based on other CpGs were also analysed: 1) using 271,785 CpGs common to the seven studies; 2) removing potential noisy probes: probes overlapping SNPs within 10 bp of the interrogated CpG, with documented SNPs at the target CpG and non-specific probes.[13] According to Illumina's annotation file, the average methylation levels across CpGs located in gene body and promoter were calculated and analysed.

Familial correlation modelling

For twin and sibling pairs whose cohabitation history was not known, we assumed for simplicity that children live with other family members from birth until age 18 years, based on evidence from a previous study.[26] In the BSGS, only 2.5% of offspring were older than 18 years, so we assumed all offspring had been living with the other family members. We assumed spouse pairs have been living

together since marriage. Under this assumption, family members in the PETS and BSGS lived together, and in the other studies lived apart (except spouse pairs) unless we had information to the contrary.

We fitted a model in which the pair correlation converges or diverges exponentially with cohabitation history. We estimated parameters ν, ω and λ $(0 \leq \omega, \nu \leq 1,$ and $\lambda \geq 0)$ for twin pairs such that

$$\rho_{ij} = \nu + \omega e^{-\lambda t} \tag{1}$$

and parameters ε, θ, λ and ν $(0 \leq \theta, \varepsilon \leq 1,$ and $\lambda, \nu \geq 0)$ for other pairs of family members such that

$$\rho_{ij} = \begin{cases} \varepsilon + \theta(1 - e^{-\lambda t}), & \text{if } t \leq t_0 \\ (\varepsilon + \theta(1 - e^{-\lambda t_0}))e^{-\nu(t-t_0)}, & \text{if } t > t_0 \end{cases} \tag{2}$$

The definition of t and t_0 depend on the relationship between i and j: 1) for twin pairs, t = age of twins; 2) for sibling pairs, t = age of the younger sibling and t_0 = age of the younger sibling when the older sibling was 18 years old; 3) for parent-offspring pairs, t = age of the offspring and t_0 = 18 years; and 4) for spouse pairs, t = time since the pair was married and t_0 = the time when the pair might have become separated (if known), where t was recorded by the KHTS, otherwise t = age of the oldest offspring of the pair in the BSGS, and the average age of the pair minus 24 years in the MCCS (the median age at first marriage for Australians during 1940s – 1970s was approximately 24 years[27]).

In equation (1), $\rho_{ij} = \nu$ when $t = \infty$, therefore ν is the correlation for twin pairs in old age, i.e. when they have lived separately for a long time. $\rho_{ij} = \nu + \omega$ when $t = 0$, therefore $\nu + \omega$ is the correlation for twin pairs at birth. Similarly, ε in equation (2) is the correlation for other family members at the beginning of cohabitation, and $\varepsilon + \theta$ in equation (2) is the correlation for other family members who have lived together for a long time. According to definition, statistical inferences for λ and ν are one-sided.

Variance component analysis

We assumed that the residual variance can be partitioned into four variance components: σ_A^2, the effects of additive genetic factors ; σ_T^2, the effects of environmental factors shared by twins alone and assumed

to be shared to the same extent within MZ and DZ pairs; σ_C^2, the effects of environmental factors shared by all family members (including twins) and assumed to be shared to the same extent within all pairs; and σ_E^2, the effects of individual-specific environmental factors and measurement error. Under such assumption, the covariance is $\sigma_A^2 + \sigma_T^2 + \sigma_C^2$ for MZ pairs, $0.5 \times \sigma_A^2 + \sigma_T^2 + \sigma_C^2$ for DZ pairs, $0.5 \times \sigma_A^2 + \sigma_C^2$ for sibling pairs and for parent-offspring pairs, and σ_C^2 for spouse pairs. According to the relationship between the familial correlation and cohabitation history, we modelled

$$\sigma_T^2 = \mu + \varphi e^{-\xi t} \tag{3}$$

in which $0 \le \varphi, \mu \le 1$ and $\xi \ge 0$, and

$$\sigma_C^2 = \begin{cases} \eta(1 - e^{-\psi t}), & \text{if } t \le t_0 \\ \eta(1 - e^{-\psi t_0})e^{-\zeta(t-t_0)}, & \text{if } t > t_0 \end{cases} \tag{4}$$

in which $0 \le \eta \le 1$ and $\psi, \zeta \ge 0$.

In this model, $\sigma_E^2 = 1 - \sigma_T^2 - \sigma_C^2 - \sigma_A^2$. According to definition, statistical inferences for ξ, ψ and ζ are one-sided.

REFERENCES

1. Saffery R, Morley R, Carlin JB, et al. Cohort profile: The peri/post-natal epigenetic twins study. *Int J Epidemiol* 2012; **41**: 55-61.
2. Powell JE, Henders AK, McRae AF, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One* 2012; **7**: e35430.
3. Sung J, Cho SI, Lee K, et al. Healthy Twin: a twin-family study of Korea--protocols and current status. *Twin Res Hum Genet* 2006; **9**: 844-8.
4. Odefrey F, Stone J, Gurrin LC, et al. Common genetic variants associated with breast cancer and mammographic density measures that predict disease. *Cancer Res* 2010; **70**: 1449-58.
5. Grundberg E, Meduri E, Sandling JK, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am J Hum Genet* 2013; **93**: 876-90.
6. Sachdev PS, Lammel A, Trollor JN, et al. A comprehensive neuropsychiatric study of elderly twins: the Older Australian Twins Study. *Twin Res Hum Genet* 2009; **12**: 573-82.
7. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC Sci Publ* 2002; **156**: 69-70.

8.   Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014; **30**: 1363-9.

9.   Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol* 2012; **13**: R44.

10.   Martino D, Loke YJ, Gordon L, et al. Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol* 2013; **14**: R42.

11.   Gordon L, Joo JE, Powell JE, et al. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Res* 2012; **22**: 1395-406.

12.   Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1215.

13.   Price ME, Cotton AM, Lam LL, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics & chromatin* 2013; **6**: 4.

14.   McRae AF, Powell JE, Henders AK, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol* 2014; **15**: R73.

15.   Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods* 2014; **11**: 1138-40.

16.   Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 2013; **29**: 189-96.

17.   Joo JE, Wong EM, Baglietto L, et al. The use of DNA from archival dried blood spots with the Infinium HumanMethylation450 array. *BMC Biotechnol* 2013; **13**: 23.

18.   Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**: 118-27.

19.   Li S, Wong EM, Joo JE, et al. Genetic and Environmental Causes of Variation in the Difference Between Biological Age Based on DNA Methylation and Chronological Age for Middle-Aged Women. *Twin Res Hum Genet* 2015; **18**: 720-6.

20.   Severi G, Southey MC, English DR, et al. Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. *Breast Cancer Res Treat* 2014; **148**: 665-73.

21.   Wong Doo N, Makalic E, Joo JE, et al. Global measures of peripheral blood-derived DNA methylation as a risk factor in the development of mature B-cell neoplasms. *Epigenomics* 2016; **8**: 55-66.

22.   Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics* 2012; **13**: 86.

23.   Eipel M, Mayer F, Arent T, et al. Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. *Aging* 2016; **8**: 1034-48.

24.   Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* 2014; **30**: 1431-9.

25.   Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics* 2011; **27**: 1496-505.

26.   Hopper JL, Green RM, Nowson CA, et al. Genetic, common environment, and individual specific components of variance for bone mineral density in 10- to 26-year-old females: a twin study. *Am J Epidemiol* 1998; **147**: 17-29.

27.   *4102.0 - Australian Social Trends, 1997*. 1997 03/22/2006 [cited 2016 May 23]; Available from:
http://www.abs.gov.au/AUSSTATS/abs@.nsf/2f762f95845417aeca25706c00834efa/a8d1bea8a2ff1b33ca2570ec001b0dc3!OpenDocument

Supplementary Table 1 Statistical analyses in each study*

Study	Number of CpGs	Methods for batch effect	Technical covariates adjusted	Covariates adjusted
PETS (birth)	330,168	Mixed effects model	Array, position on the array	Sex
PETS (18 months)	330,168	Mixed effects model	Array, position on the array	Sex[†]
BSGS	417,069	Mixed effects model	Array, position on the array	Age, sex [‡]
KHTS	459,805	Mixed effects model	Array, position on the array	Age, sex, smoking, BMI, experiment (I and II)
AMDTSS	468,406	*ComBat*	–	Age
MuTHER	460,832	Mixed effects model	Array	Age
OATS	444,330	Mixed effects model	Array, position on the array	Age, sex, DNA extraction protocol, experiment (I and II), recruited state
MCCS	473,482	*ComBat*	–	Age, sex, sub-study, sample type

* PETS, Peri/postnatal Epigenetic Twins Study; BSGS, Brisbane Systems Genetic Study; KHTS, Korean Healthy Twin Study; AMDTSS, Australian Mammographic Twins and Sisters Study; MuTHER, Multiple Tissue Human Expression Resource Study; OATS, Older Australian Twins Study; MCCS, Melbourne Collaborative Cohort Study; MZ, monozygotic; DZ, dizygotic; BMI, body mass index.

† Adjustment stratified by zygosity, for that there was a difference in the mean of genome-wide average DNA methylation between monozygotic and dizygotic twins.

‡ Adjustment stratified by generation, for that there was a difference in the mean of genome-wide average DNA methylation between generations.

Supplementary Table 2 Number of pairs of family members in each study*

Pairs	PETS (birth)	PETS (18 months)	BSGS	KHTS	AMDTSS	MuTHER	OATS	MCCS
MZ pairs	9	6	67	91	66	93	108	–
DZ pairs	5	4	111	–	66	153	–	–
Sibling pairs	–	–	260	151	552	–	–	–
Parent-offspring pairs	–	–	363	321	–	–	–	–
Spouse pairs	–	–	59	69	–	–	–	43

* PETS, Peri/postnatal Epigenetic Twins Study; BSGS, Brisbane Systems Genetic Study; KHTS, Korean Healthy Twin Study; AMDTSS, Australian Mammographic Twins and Sisters Study; MuTHER, Multiple Tissue Human Expression Resource Study; OATS, Older Australian Twins Study; MCCS, Melbourne Collaborative Cohort Study; MZ, monozygotic; DZ, dizygotic.

Supplementary Table 3 Familial correlation estimates in genome-wide average DNA methylation additionally adjusted for cell mixture in each study*

Pairs	PETS (birth)	PETS (18 months)	BSGS	KHTS	AMDTSS	OATS	MCCS
MZ pairs	0.65	0.78	0.57	0.4	0.44	0.30	–
	(0.43, 0.80)	(0.68, 0.86)	(0.46, 0.66)	(0.24, 0.54)	(0.27, 0.58)	(0.14, 0.44)	
DZ pairs	0.81	0.88	0.36	–	0.43	–	–
	(0.70, 0.88)	(0.84, 0.92)	(0.23, 0.49)		(0.27, 0.56)		
Twin pairs combined	0.70	0.82	0.43	–	0.43	–	–
	(0.58, 0.80)	(0.76, 0.87)	(0.34, 0.52)		(0.32, 0.53)		
Sibling pairs			0.26	0.28	−0.01	–	–
			(0.13, 0.37)	(0.09, 0.45)	(−0.12, 0.10)		
Parent-offspring pairs			0.24	0.17	–	–	–
			(0.14, 0.34)	(0.04, 0.29)			
Spouse pairs			0.17	0.24	–	–	0.28
			(−0.03, 0.36)	(0.05, 0.42)			(0.02, 0.50)

* Results are presented as estimate (95% confidence interval). PETS, Peri/postnatal Epigenetic Twins Study; BSGS, Brisbane Systems Genetic Study; KHTS, Korean Healthy Twin Study; AMDTSS, Australian Mammographic Twins and Sisters Study; OATS, Older Australian Twins Study; MCCS, Melbourne Collaborative Cohort Study; MZ, monozygotic; DZ, dizygotic.

Supplementary Table 4 Familial correlation estimates in genome-wide average DNA methylation using the 271,785 common CpGs in each study*

Pairs	PETS (birth)	PETS (18 months)	BSGS	KHTS	AMDTSS	MuTHER	OATS	MCCS
MZ pairs	0.77 (0.67, 0.84)	0.78 (0.67, 0.86)	0.56 (0.44, 0.65)	0.31 (0.14, 0.45)	0.34 (0.14, 0.51)	0.25 (0.09, 0.40)	0.32 (0.16, 0.45)	–
DZ pairs	0.82 (0.74, 0.88)	0.89 (0.85, 0.92)	0.40 (0.27, 0.50)	–	0.40 (0.24, 0.54)	0.29 (0.15, 0.42)	–	–
Twin pairs combined	0.79 (0.72, 0.84)	0.82 (0.75, 0.86)	0.45 (0.36, 0.53)	–	0.37 (0.25, 0.48)	0.28 (0.17, 0.37)	–	–
Sibling pairs	–	–	0.26 (0.13, 0.38)	0.18 (−0.08, 0.41)	0.05 (−0.06, 0.15)	–	–	–
Parent-offspring pairs	–	–	0.26 (0.15, 0.36)	0.18 (0.04, 0.30)	–	–	–	–
Spouse pairs	–	–	0.35 (0.16, 0.52)	0.24 (0.05, 0.42)	–	–	–	0.30 (0.04, 0.51)

* Results are presented as estimate (95% confidence interval). PETS, Peri/postnatal Epigenetic Twins Study; BSGS, Brisbane Systems Genetic Study; KHTS, Korean Healthy Twin Study; AMDTSS, Australian Mammographic Twins and Sisters Study; MuTHER, Multiple Tissue Human Expression Resource Study; OATS, Older Australian Twins Study; MCCS, Melbourne Collaborative Cohort Study; MZ, monozygotic; DZ, dizygotic.

Supplementary Table 5 Familial correlation estimates in genome-wide average DNA methylation without potential noisy CpGs in each study*

Pairs	PETS (birth)	PETS (18 months)	BSGS	KHTS	AMDTSS	MuTHER	OATS	MCCS
MZ pairs	0.76	0.75	0.57	0.40	0.39	0.30	0.34	–
	(0.66, 0.84)	(0.61, 0.84)	(0.47, 0.66)	(0.25, 0.54)	(0.22, 0.54)	(0.14, 0.45)	(0.19, 0.47)	
DZ pairs	0.83	0.89	0.40	–	0.36	0.47	–	–
	(0.75, 0.88)	(0.86, 0.92)	(0.28, 0.51)		(0.18, 0.51)	(0.38, 0.55)		
Twin pairs combined	0.79	0.80	0.46	–	0.37	0.40	–	–
	(0.72, 0.84)	(0.73, 0.85)	(0.37, 0.53)		(0.25, 0.48)	(0.32, 0.48)		
Sibling pairs	–	–	0.28	0.26	0.02	–	–	–
			(0.15, 0.39)	(0.05, 0.44)	(−0.09, 0.13)			
Parent-offspring pairs	–	–	0.25	0.17	–	–	–	–
			(0.15, 0.35)	(0.04, 0.29)				
Spouse pairs	–	–	0.27	0.21	–	–	–	0.31
			(0.04, 0.46)	(0.01, 0.39)				(0.06, 0.52)

* Results are presented as estimate (95% confidence interval). PETS, Peri/postnatal Epigenetic Twins Study; BSGS, Brisbane Systems Genetic Study; KHTS, Korean Healthy Twin Study; AMDTSS, Australian Mammographic Twins and Sisters Study; MuTHER, Multiple Tissue Human Expression Resource Study; OATS, Older Australian Twins Study; MCCS, Melbourne Collaborative Cohort Study; MZ, monozygotic; DZ, dizygotic.

Supplementary Table 6 Familial correlation estimates in gene body average DNA methylation in each study*

Pairs	PETS (birth)	PETS (18 months)	BSGS	KHTS	AMDTSS	MuTHER	OATS	MCCS
MZ pairs	0.88	0.91	0.57	0.46	0.49	0.24	0.28	–
	(0.85, 0.90)	(0.89, 0.93)	(0.47, 0.66)	(0.32, 0.58)	(0.36, 0.61)	(0.06, 0.41)	(0.12, 0.43)	
DZ pairs	0.83	0.87	0.36	–	0.41	0.28	–	–
	(0.76, 0.89)	(0.81, 0.91)	(0.24, 0.48)		(0.23, 0.55)	(0.14, 0.40)		
Twin pairs combined	0.86	0.89	0.43	–	0.45	0.27	–	–
	(0.84, 0.89)	(0.87, 0.91)	(0.34, 0.51)		(0.35, 0.55)	(0.16, 0.37)		
Sibling pairs	–	–	0.25	0.32	-0.02	–	–	–
			(0.11, 0.37)	(0.12, 0.49)	(−0.13, 0.09)			
Parent-offspring pairs	–	–	0.26	0.19	–	–	–	–
			(0.16, 0.35)	(0.06, 0.30)				
Spouse pairs	–	–	0.22	0.33	–	–	–	0.21
			(−0.01, 0.42)	(0.15, 0.49)				(−0.08, 0.46)

* Results are presented as estimate (95% confidence interval). PETS, Peri/postnatal Epigenetic Twins Study; BSGS, Brisbane Systems Genetic Study; KHTS, Korean Healthy Twin Study; AMDTSS, Australian Mammographic Twins and Sisters Study; MuTHER, Multiple Tissue Human Expression Resource Study; OATS, Older Australian Twins Study; MCCS, Melbourne Collaborative Cohort Study; MZ, monozygotic; DZ, dizygotic.

Supplementary Table 7 Familial correlation estimates in gene promoter average DNA methylation in each study*

Pairs	PETS (birth)	PETS (18 months)	BSGS	KHTS	AMDTSS	MuTHER	OATS	MCCS
MZ pairs	0.35	0.09	0.47	0.28	0.21	0.01	0.41	–
	(−0.16, 0.71)	(−0.51, 0.63)	(0.32, 0.60)	(0.11, 0.44)	(−0.02, 0.42)	(−0.18, 0.21)	(0.28, 0.52)	
DZ pairs	0.80	0.86	0.39	–	0.30	0.08	–	–
	(0.69, 0.87)	(0.80, 0.91)	(0.26, 0.50)		(0.11, 0.46)	(−0.07, 0.24)		
Twin pairs combined	0.50	0.25	0.42	–	0.26	0.06	–	–
	(0.18, 0.72)	(−0.32, 0.69)	(0.32, 0.51)		(0.11, 0.39)	(−0.07, 0.18)		
Sibling pairs	–	–	0.24	0.08	0.05	–	–	–
			(0.10, 0.36)	(−0.14, 0.31)	(−0.05, 0.16)			
Parent-offspring pairs	–	–	0.18	0.13	–	–	–	–
			(0.08, 0.29)	(0.01, 0.25)				
Spouse pairs	–	–	0.24	0.11	–	–	–	0.16
			(0.03, 0.43)	(−0.11, 0.31)				(−0.13, 0.42)

* Results are presented as estimate (95% confidence interval). PETS, Peri/postnatal Epigenetic Twins Study; BSGS, Brisbane Systems Genetic Study; KHTS, Korean Healthy Twin Study; AMDTSS, Australian Mammographic Twins and Sisters Study; MuTHER, Multiple Tissue Human Expression Resource Study; OATS, Older Australian Twins Study; MCCS, Melbourne Collaborative Cohort Study; MZ, monozygotic; DZ, dizygotic.

Supplementary Table 8 Parameter estimates for the relationship between correlation and cohabitation history from modelling the combined data of the seven studies

Paris of family members	Parameter	Estimate	Standard Error
Twin pairs	ν	0.37	0.04
	ω	0.48	0.06
	$\lambda_{twin}^{\dagger}$	1.15	0.43
Pairs of non-twin first-degree relatives	ε_{1st}	0.00	–
	θ_{1st}	0.75	0.29
	$\lambda_{1st}^{\dagger}$	0.32	0.18
	$\nu^{\dagger}$	0.64	0.36
Spouse pairs	ε_{spouse}	0.00	–
	θ_{spouse}	0.34	0.16
	$\lambda_{spouse}^{\dagger}$	0.83	0.42

† Reported as the change per 10 years.

Supplementary Table 9 Parameter estimates from the variance components model for the combined data of the seven studies*

Parameter	Estimate	Standard Error
μ	0.28	0.06
φ	0.63	0.08
ξ	0.24	0.11
η	0.29	0.10
ψ	0.13	0.14
ζ	0.02	0.01
σ_A^2	−0.07	0.09
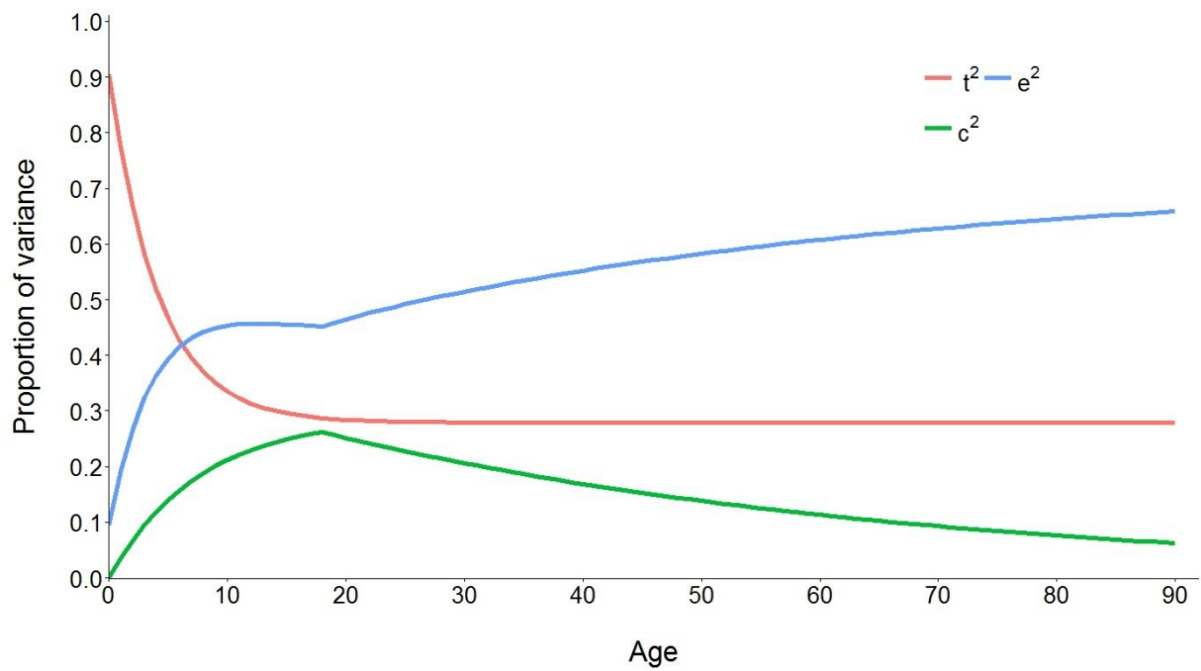
Supplementary Figure 1 Proportion of variance explained by each variance component estimated from the variance components model for the combined data from the seven studies

t^2 represents the effects of environmental factors shared by twins alone, c^2 represents the effects of environmental factors shared by all family members (including twins), and e^2 represents the effects of individual-specific environmental factors and measurement error. Since the effects of additive genetic factors were estimated to be negative, they are not reported in the figure.