



Universiteit  
Leiden  
The Netherlands

## **Epigenome-wide change and variation in DNA methylation in childhood: trajectories from birth to late adolescence**

Mulder, R.H.; Neumann, A.; Cecil, C.A.M.; Walton, E.; Houtepen, L.C.; Simpkin, A.J.; ... ; Suderman, M.

### **Citation**

Mulder, R. H., Neumann, A., Cecil, C. A. M., Walton, E., Houtepen, L. C., Simpkin, A. J., ... Suderman, M. (2021). Epigenome-wide change and variation in DNA methylation in childhood: trajectories from birth to late adolescence. *Human Molecular Genetics*, 30(1), 119-134. doi:10.1093/hmg/ddaa280

Version: Publisher's Version  
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)  
Downloaded from: <https://hdl.handle.net/1887/3204176>

**Note:** To cite this publication please use the final published version (if applicable).

ASSOCIATION STUDIES ARTICLE

# Epigenome-wide change and variation in DNA methylation in childhood: trajectories from birth to late adolescence

Rosa H. Mulder<sup>1,2,3,†</sup>, Alexander Neumann<sup>1,2,4,§</sup>, Charlotte A.M. Cecil<sup>1,5,6,§</sup>, Esther Walton<sup>7,8,†</sup>, Lotte C. Houtepen<sup>7</sup>, Andrew J. Simpkin<sup>7,9</sup>, Jolien Rijlaarsdam<sup>1,2</sup>, Bastiaan T. Heijmans<sup>10</sup>, Tom R. Gaunt<sup>7</sup>, Janine F. Felix<sup>2,11</sup>, Vincent W.V. Jaddoe<sup>2,11</sup>, Marian J. Bakermans-Kranenburg<sup>12</sup>, Henning Tiemeier<sup>1,13</sup>, Caroline L. Relton<sup>7</sup>, Marinus H. van IJzendoorn<sup>14,15</sup> and Matthew Suderman<sup>7,\*</sup>

<sup>1</sup>Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands, <sup>2</sup>Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands, <sup>3</sup>Institute of Education and Child Studies, Leiden University, Leiden, The Netherlands, <sup>4</sup>Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada, <sup>5</sup>Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands, <sup>6</sup>Department of Psychology, Institute of Psychology, Psychiatry and Neuroscience, King's College London, London, UK, <sup>7</sup>MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK, <sup>8</sup>Department of Psychology, University of Bath, Bath, UK, <sup>9</sup>School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland, <sup>10</sup>Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands, <sup>11</sup>Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands, <sup>12</sup>Clinical Child and Family Studies, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, <sup>13</sup>Department of Social and Behavioral Science, Harvard T.H. Chan School of Public Health, Boston, MA, USA, <sup>14</sup>Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands and <sup>15</sup>School of Clinical Medicine, University of Cambridge, Cambridge, UK

\*To whom correspondence should be addressed at: University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. Tel: +44 (0)117 331 0090; Email: matthew.suderman@bristol.ac.uk

## Abstract

DNA methylation (DNAm) is known to play a pivotal role in childhood health and development, but a comprehensive characterization of genome-wide DNAm trajectories across this age period is currently lacking. We have therefore

<sup>†</sup>Rosa H. Mulder, <http://orcid.org/0000-0002-2382-1704>

<sup>‡</sup>Esther Walton, <http://orcid.org/0000-0002-0935-2200>

<sup>§</sup>Both authors have contributed equally to this work.

Received: October 8, 2020. Revised: December 3, 2020. Accepted: December 23, 2020

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

performed a series of epigenome-wide association studies in 5019 blood samples collected at multiple time-points from birth to late adolescence from 2348 participants of two large independent cohorts. DNAm profiles of autosomal CpG sites (CpGs) were generated using the Illumina Infinium HumanMethylation450 BeadChip. Change over time was widespread, observed at over one-half (53%) of CpGs. In most cases, DNAm was decreasing (36% of CpGs). Inter-individual variation in linear trajectories was similarly widespread (27% of CpGs). Evidence for non-linear change and inter-individual variation in non-linear trajectories was somewhat less common (11 and 8% of CpGs, respectively). Very little inter-individual variation in change was explained by sex differences (0.4% of CpGs) even though sex-specific DNAm was observed at 5% of CpGs. DNAm trajectories were distributed non-randomly across the genome. For example, CpGs with decreasing DNAm were enriched in gene bodies and enhancers and were annotated to genes enriched in immune-developmental functions. In contrast, CpGs with increasing DNAm were enriched in promoter regions and annotated to genes enriched in neurodevelopmental functions. These findings depict a methylome undergoing widespread and often non-linear change throughout childhood. They support a developmental role for DNA methylation that extends beyond birth into late adolescence and has implications for understanding life-long health and disease. DNAm trajectories can be visualized at <http://epidelta.mrcieu.ac.uk>.

## Introduction

DNA methylation (DNAm), an epigenetic process whereby DNA is modified by the addition of methyl groups, has gained increasing attention over the past few decades, due to its pivotal role in development. *In utero*, DNAm is involved in a range of essential processes, including cell differentiation (1–3), X-chromosome inactivation (4) and fetal growth (5). Its role extends well beyond birth, e.g. by maintaining cell type identity and genome stability (6–8), responding to environmental exposures (9–11), and its involvement, among many other processes, in immune (12) and neural development (13). Since it is influenced by both genetic and environmental factors (14,15), DNAm has also emerged as a key mechanism of interest for understanding the gene-environmental interplay in normal ageing and disease development.

Numerous studies have identified strong associations between DNAm and age. While most have relied on cross-sectional data (16–18), a few have utilized longitudinal measurements of DNAm within individuals (19–23). Longitudinal measurements allow one to distinguish intra-individual change from inter-individual differences in change, thereby greatly improving the power to detect change over time and to identify differences between individuals (24). Identifying and characterizing CpGs for which DNAm changes differently over time between individuals (i.e. inter-individual variation in change) is a necessary step in identifying genetic and environmental influences on the methylome as well as their potential impact on health outcomes (25). Moreover, longitudinal designs facilitate the study of non-linear trajectories (26,27), which might help to identify sensitive periods for DNAm change in development. To date, the largest epigenome-wide longitudinal study on DNAm included 385 elderly individuals who were followed up to five times over a maximum period of 18 years, identifying DNAm change at 1316 CpG (Cytosine-phosphate-Guanine) sites (19) and change of inter-individual variation at 570 CpGs (20). Yet, little is known about DNAm trajectories across early development, as existing studies in childhood DNAm typically have been limited by small sample sizes (21,23), short time-periods (22,28) or focused on specific CpGs in relation to maternal smoking (29), birthweight (30) or maternal BMI (31).

In the current study, we aim to provide a benchmark of typical epigenome-wide age-related DNAm trajectories within individuals, spanning the first two decades of life. This study combines repeated measurements of DNAm in blood at nearly half a million CpG sites across the genome from two large population-based cohorts, the Generation R Study (Generation R)

and Avon Longitudinal Study of Parents and Children (ALSPAC), to form one integrated dataset with four time-points of measurement. In a series of three epigenome-wide mixed model analyses, we study linear (Model 1), non-linear (Model 2) and sex-related (Model 3) trajectories of change across development. Furthermore, we aim to identify CpGs for which trajectories vary between individuals (Models 1 and 2). Results are interpreted in the context of CpG location and biological pathways. The key findings are discussed here; full results per CpG can be freely accessed and visualized at <http://epidelta.mrcieu.ac.uk/>.

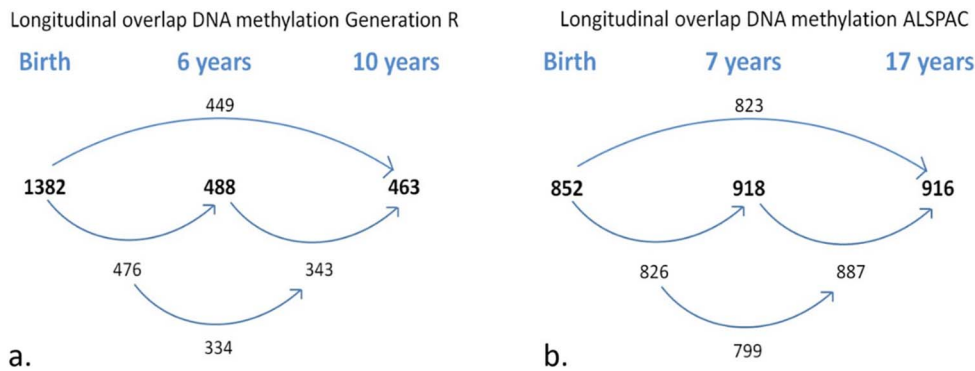
## Results

### Cross-cohort comparability

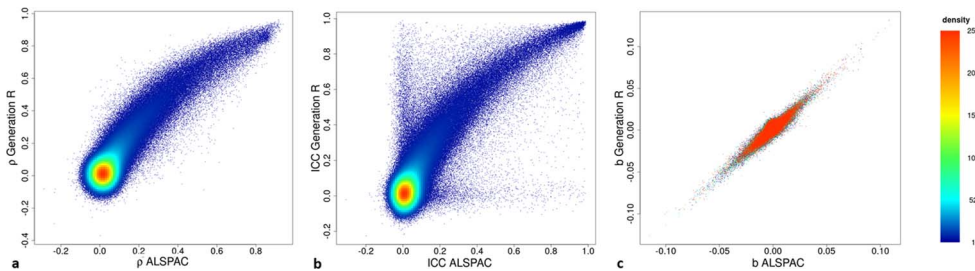
Sample characteristics of 1399 Generation R participants (total DNAm samples = 2333) and of 949 ALSPAC participants (total DNAm samples = 2686; Fig. 1) are provided in [Supplementary Material, Table S1](#). After the DNAm datasets of the two cohorts underwent joint functional normalization (see [Supplementary Material, Fig. S1](#) for distributions of mean DNAm levels), within-cohort stability of DNAm at birth and 6 or 7 years (in Generation R and ALSPAC, respectively) was compared. Stability of DNAm at individual CpG sites (437 864 autosomal sites) was estimated in three ways: relative concordance using Spearman correlations between time points, absolute concordance using intra-class correlations between time points (children with data for both time points:  $n$  Generation R = 476,  $n$  ALSPAC = 826) and change over time using change estimates from a linear mixed model (Model 1, see Materials and Methods) applied within each cohort (children with data for at least one of the two time-points:  $n$  Generation R = 1394,  $n$  ALSPAC = 944). Estimates of all stability measures for both cohorts are depicted in [Fig. 2](#). Next, agreement of these stability estimates between the two cohorts was estimated with the Spearman ( $\rho$ ) or Pearson ( $r$ ) correlation (depending on normality of the data) across all CpGs, between the datasets. The Spearman correlation of the relative concordance was  $\rho = 0.62$ , the Pearson correlation of the absolute concordance was  $\rho = 0.60$  and the Pearson correlation of the change estimates was  $r = 0.86$ , indicating strong agreement between datasets. Based on these results, the two datasets were joined to form one set with four different time-points of DNAm (birth, age 6/7, 10 and 17 years).

### Linear DNAm change from birth to early adulthood

Estimates of overall change in DNAm from birth to early adolescence (Model 1; see Materials and Methods) indicated



**Figure 1.** Longitudinal sample sizes for (a) Generation R ( $N$  total children = 1399,  $N$  total DNAm samples = 2333) and (b) ALSPAC ( $N$  total children = 949,  $N$  total DNAm samples = 2686). Bolded numbers represent total sample size at each time-point; non-bolded number refers to overlapping samples between time-points.



**Figure 2.** Scatterplots of within-cohort stability of DNA methylation showing (a) Spearman correlations, (b) intra-class correlation coefficients and (c) change estimates from birth to 6/7 years per CpG for Generation R and ALSPAC.

linear change at 51.6% of CpGs at a Bonferroni-corrected threshold ( $P < 1 \times 10^{-7}$ ) (Fig. 3a and b). Specifically, DNAm decreased over time at 35.5% of all CpGs and increased at 16.0% (Fig. 4). The mode intercept indicated that the decreasing CpGs were 88% methylated at birth ( $SD = 32\%$ ) (see Fig. 5 for the distribution of intercept for different groups of CpGs). DNAm levels for increasing CpGs typically started at 5% ( $SD = 23\%$ ).

The mode estimate of the trimodally distributed (Supplementary Material, Figure S2) DNAm change was  $b = -9.24 \times 10^{-4}$  (with corresponding mode  $SE = 6.85 \times 10^{-5}$ ), indicating an overall 0.09% DNAm decrease per year at a typical CpG site. This translates into a 1.66% decrease in DNAm over the course of 18 years. An example of a CpG site with a typical change in DNAm is depicted in Fig. 3a. The largest observed absolute change in DNAm was  $b = -3.47 \times 10^{-2}$  ( $SE = 3.65 \times 10^{-4}$ ,  $P < 9.88 \times 10^{-324}$ ), indicating an overall DNAm decrease of 62.5% over 18 years (Fig. 3b). Only 22 CpGs showed an absolute change  $>50\%$  over the course of 18 years (Supplementary Material, Table S2). These CpGs were mainly annotated to TSS200 regions, the gene body or intergenic regions and annotated to genes associated with global biological functions, such as protein-protein interactions [TTC22, (32)] or transcriptional regulation [NFIX, (33)] as well as more specific functions, such as synaptic scaffolding [SHANK2, (34)] and muscle regulation [CSR3, (35)]. Associated diseases include autism spectrum disorder [SHANK2, (36)], cardiomyopathy [CSR3, (37)] and Malan syndrome, with the latter also being characterized as an overgrowth disorder (38). With such a low number of sites that display such a large change, it follows that typically in (cord-/peripheral) blood tissue, DNAm levels for CpGs do not change from a fully unmethylated to fully methylated state, or vice versa, over the course of 18 years.

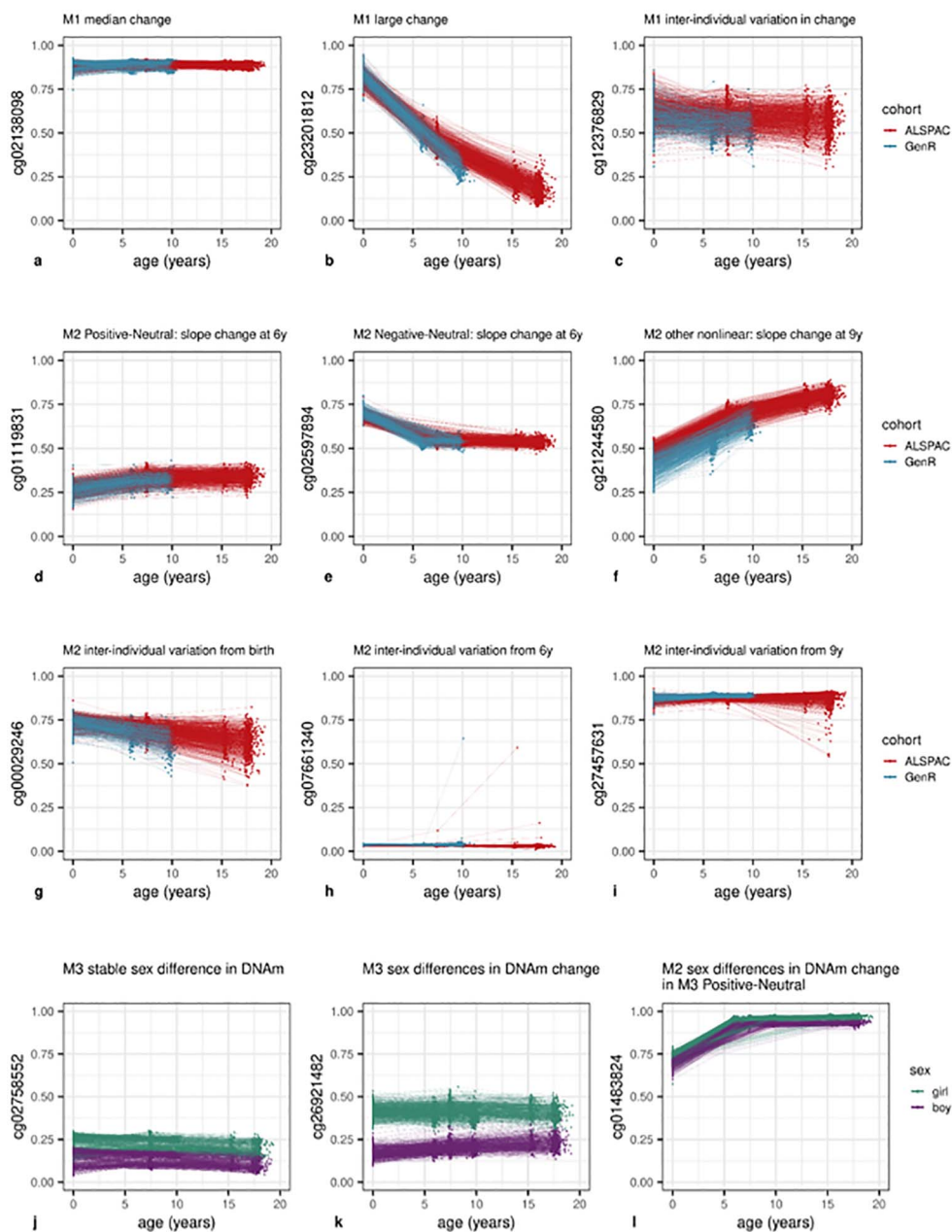
Furthermore, we observed substantial inter-individual variation in linear DNAm changes over time at 27.4% of all

CpGs (i.e. random slope variance was greater than zero at Bonferroni-corrected threshold  $P < 1 \times 10^{-7}$ ; Fig. 3c). On average, this variation accounted for 2.7% ( $SD = 1.5\%$ ) of all estimated inter-individual variation (for intercept, age, batch and residual; see also Supplementary Material, Table S3A) at these CpGs. At 17.3% of all CpGs, we observed both change and inter-individual variation in change.

### Nonlinear DNAm change

Model 2 (see Materials and Methods) was identical to Model 1, but permitted slope changes at ages 6 and 9 years to test for non-linear DNAm trajectories. At 11.0% of CpGs, a non-linear trajectory was detected. Specifically, at 4.8% of all CpGs, DNAm increased from birth and remained stable from 6 onward (Positive-Neutral; Fig. 3d). Second, at 3.1% of all CpGs, DNAm decreased from birth and then remained stable at 6 years (Negative-Neutral; Fig. 3e). The remaining 3.0% of all CpGs followed other non-linear trajectories (e.g. Fig. 3f), with each trajectory observed in  $<1.0\%$  of all CpGs. Overall, linear and/or non-linear changes in Models 1 or 2 were observed in 52.6% of CpGs (Fig. 3), indicating that most non-linear patterns were also detected as linear patterns in Model 1.

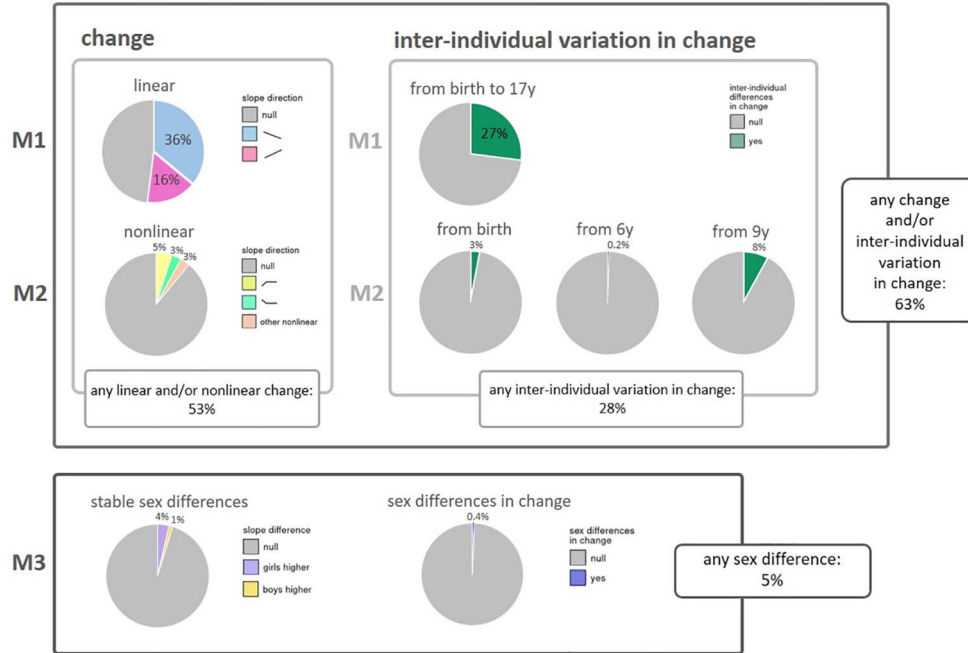
Inter-individual differences in change (i.e. random variance in slopes) from birth onward were detected at 3.4% of all sites (Fig. 3g), inter-individual differences in slope change at 6 years in 0.2% (Fig. 3h) and inter-individual differences in slope change at 9 years at 8.2% of CpGs (Fig. 3i). For CpGs at which inter-individual differences in change from birth onwards were detected, this variation on average accounted for 3.9% ( $SD = 9.3\%$ ) of all inter-individual variation, for CpGs at which inter-individual differences in slope change at 6 years were detected, this variation on average accounted for 20.6% ( $SD = 13.2\%$ ) of all



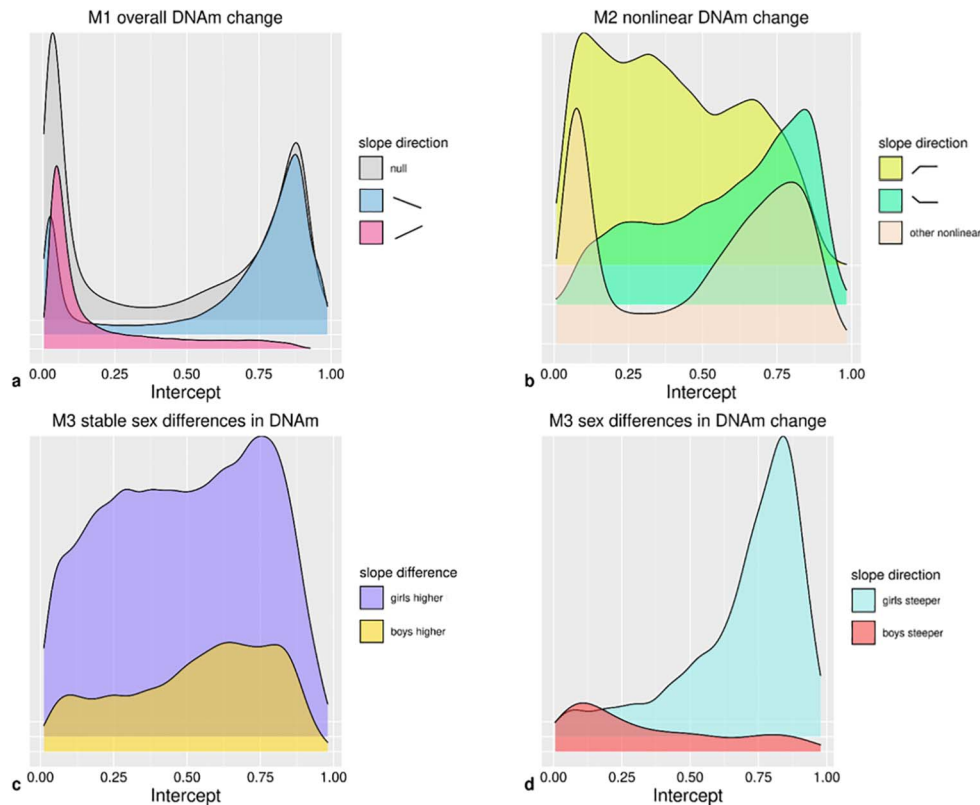
**Figure 3.** DNAm levels of selected CpG sites across childhood. Parts (a–c) show CpG sites with linear change over time (Model 1). A typical site is shown in (a), the site with the largest observed change in (b) and with inter-individual variation in DNAm change (c). Parts (d–f) show CpG sites with non-linear change (Model 2). A Positive-Neutral trajectory is shown in (d), a Negative-Neutral trajectory in (e) and a Positive-More Positive-Less Positive in (f). Parts (g–i) show CpG sites with inter-individual variation in change (Model 2). A site with slope variation from birth is shown in (g), slope change variation at 6 in (h) and slope change variation at 9 in (i). Parts (j–l) show CpG sites with sex-specific DNAm. A site with stable sex differences is shown in (Model 3) (j), sex-specific slope in (Model 3) (k) and sex-specific slope change at 6 in (Model 2) (l).

inter-individual variation and for CpGs at which differences in slope change at 9 years were detected, this variation on average accounted for 23.3% ( $SD = 17.6\%$ ) of all inter-individual variation (for a complete overview of inter-individual variation within all CpGs, see [Supplementary Material, Table S3B](#)). We note that CpGs with inter-individual differences in change at 6 years often contained a few extreme outliers. These outliers, however, were found across cohorts and batches and were often found within individuals across time-points, indicating that these outliers seem to reflect true DNAm values rather than technical issues.

Inter-individual differences in slope (change) at each time-point were detected more often at CpGs with an increasing rather than decreasing overall DNAm change in Model 1 ( $P \leq 2.37 \times 10^{-144}$ ). At last, both Positive-Neutral and Negative-Neutral changes coincided more often with inter-individual variation from birth ( $P < 9.88 \times 10^{-324}$ ). Any inter-individual differences in change, detected by Models 1 or 2, were observed at 27.9% of CpGs. In total, Models 1 and 2 detected age-related change whether linear, non-linear or inter-individual differences in change at 62.8% of all CpG sites ([Fig. 3](#)).



**Figure 4.** Overview of results from the three models. Model 1 (M1) was applied for overall change in DNA methylation and inter-individual variation in linear change; Model 2 (M2) for non-linear change in DNA methylation and inter-individual variation in non-linear change; and Model 3 (M3) for stable sex differences in DNA methylation and sex differences in change of DNA methylation (sex by time interaction). Percentages represent percentage of autosomal CpGs below Bonferroni-corrected threshold ( $P < 1 \times 10^{-7}$ ).



**Figure 5.** Density plots of intercepts for CpGs with (a) directions of change in Model 1 ( $n=473\,864$ ); (b) non-linear trajectories in Model 2 ( $n=52\,043$ ); (c) stable sex differences in Model 3 ( $n=22\,821$ ); (d) sex differences in DNAm change in Model 3 ( $n=1768$ ).

## Sex differences in longitudinal DNAm and DNAm change

According to Model 3 (see Materials and Methods), sex differences in DNAm were present at 4.9% of (autosomal) CpGs (Fig. 3). Specifically, stable longitudinal sex differences (main sex effects) were observed at 4.8% of all (autosomal) CpGs (Fig. 3j), and sex differences in DNAm change (sex by age interaction effects) were found at 0.4% of all (autosomal) CpGs (Fig. 3k). At sites with stable sex differences, DNAm levels were higher in girls at 3.6% (Fig. 3j) and lower at 1.2% of CpG sites. DNAm at sites with higher DNAm in girls tended to increase over time, whereas DNAm at sites with higher DNAm in boys tended to decrease ( $P = 4.20 \times 10^{-205}$ ). Most commonly (at 0.2% of all CpGs), DNAm was higher in girls at birth but DNAm in boys increased at a higher rate.

Both CpGs with stable sex differences and those with sex differences in DNAm change were less likely to show inter-individual variation in DNAm change than other sites (20.8 versus 27.5% and 18.1 versus 27.3%;  $P = 5.36 \times 10^{-111}$  and  $P = 7.57 \times 10^{-18}$ ). Finally, CpGs with stable sex differences or sex differences in DNAm change detected in Model 3 were much more likely to follow an overall Positive-Neutral trajectory of DNAm change detected in Model 2 than other CpG sites were (24.2% of CpGs with stable sex differences followed a Positive-Neutral trajectory versus 3.8% of other CpGs and 53.9% of CpGs with sex differences in DNAm change followed a Positive-Neutral trajectory versus 4.6% of other CpGs;  $P < 9.88 \times 10^{-324}$ ,  $P < 9.88 \times 10^{-324}$ ; Fig. 3l). Albeit less prominently so, CpGs with stable sex differences or sex differences in DNAm change also more often followed a Negative-Neutral trajectory than other CpGs did (stable sex differences: 5.0 versus 3.0%,  $P = 5.43 \times 10^{-62}$ ; sex differences in DNAm change: 7.7 versus 3.1%,  $P < 7.11 \times 10^{-28}$ ).

## Follow-up analyses

Follow-up analyses were performed to understand how different types of age-related DNAm trajectories are distributed across the genome (Supplementary Material, Tables S4–S6) and in association with genetic variance (Supplementary Material, Tables S7–S9). Additionally, to understand how different types of DNAm change are represented in epigenetic age estimators, we studied their enrichment in Horvath's (39) and Hannum's (40) age estimators (Supplementary Material, Tables S10 and S11). All reported enrichments have significance below a Bonferroni-corrected threshold of  $P < 3.29 \times 10^{-4}$ , corrected for the number of chi-square tests ( $n = 152$ ). We further report the enrichment of Gene Ontology (GO) pathways (nominal  $P < 0.05$ ) for genes annotated to CpG sites in each trajectory (Supplementary Material, Tables S12–S14). At last, we study the enrichment of age-related DNAm trajectories in reported hits of different epigenome-wide association studies (EWASs) (Fig. 6). All reported EWAS enrichments are below a Bonferroni-corrected threshold of  $P < 1.38 \times 10^{-04}$ , corrected for the number of Fishers' exact tests ( $n = 363$ ; Supplementary Material, Table S15).

**Patterns of DNAm change and CpG location.** CpG sites with DNAm change associated patterns were labelled by gene associated regions, CpG island associated regions as well as enhancer elements. Although many exceptions exist, low levels of DNAm in the promoter area but high levels of DNAm in the gene body are generally associated with increased gene transcription (41,42). CpGs annotated to TSS200 regions more often showed an overall DNAm increase (Model 1) than other CpGs (19.0 versus 15.6%), whereas CpGs annotated to the gene body more often showed an overall DNAm decrease than

other sites (38.8 versus 33.7%). TSS200 CpGs showed less inter-individual variation in overall DNAm change than other sites (22.2 versus 28.1%), whereas gene body CpGs showed somewhat more inter-individual variation in overall DNAm change than other sites (28.9 versus 26.5%).

Promoter areas often coincide with CpG islands (43). Here, 63.3% of TSS200 CpGs were also annotated to CpG islands. As in TSS200 areas, CpGs annotated to CpG islands had lower DNAm levels [mode M1 intercept 2.4% ( $SD = 30.2\%$ )], and more often showed an overall DNAm increase than other sites (25.2 versus 12.0%). DNAm sex differences were especially present in the shores of CpG islands compared with all other island associated regions (stable sex differences: 7.5 versus 4.0%, sex differences in DNAm change: 0.6 versus 0.3%).

Enhancers act on promoters to regulate gene transcription (44). CpGs annotated to enhancer elements (2.0% of CpGs) tended to have low DNAm levels (mode M1 intercept 5.07%;  $SD = 31.4\%$ ) and then increased with age more than other CpGs (23.9 versus 15.9%). Inter-individual variation in change from birth was more common at enhancer sites than at other sites (5.6 versus 3.3%).

**Genetic associations with patterns of DNAm change.** We looked up methylation quantitative trait loci in *cis* (*cis* meQTLs) and in *trans* (*trans* meQTLs) (45) and CpGs located on a single nucleotide polymorphism (polymorphic CpGs) (46) to investigate possible relationships between genetic variation and DNAm trajectories and sex differences. CpGs with *cis* (30.0% of CpGs) and *trans* (2.1% of CpGs) meQTLs were more likely to have increasing DNAm than other sites (*cis*: 23.4 versus 12.9% at other sites; *trans*: 27.5 versus 15.8%) and to exhibit more inter-individual variation in overall DNAm change (*cis*: 30.8 versus 25.9%; *trans*: 29.4 versus 27.3%). This inter-individual variation tended to appear from birth onwards (*cis*: 5.4 versus 2.5%; *trans*: 13.3 versus 3.1%) and not from 9 years onwards (*cis*: 7.7 versus 8.4%; *trans*: 4.7 versus 8.2%). CpG sites with meQTLs were more likely to exhibit stable DNAm sex differences (*cis*: 9.0 versus 3.0%; *trans*: 15.2 versus 4.6%) and sex differences in DNAm change than other sites (*cis*: 0.7 versus 0.2%; *trans*: 2.0 versus 0.3%).

Polymorphic CpGs (or for Infinium I, single nucleotide polymorphisms at site of single base extension) (14.0% of all CpGs) more often had decreasing (42.72 versus 34.35%) rather than increasing DNAm (9.16 versus 17.15%) than other CpGs had. Furthermore, there was less inter-individual variation in the slope change at 9 years (5.86 versus 8.54%) and less stable sex differences were found at polymorphic CpGs (3.61 versus 5.01%).

In sum, CpGs linked to meQTLs more often had increasing DNAm and DNAm sex differences, whereas polymorphic CpGs less often had increasing DNAm and DNAm sex differences. Both CpGs linked to meQTLs and polymorphic CpGs less often showed inter-individual variation in the slope change at 9 years.

**Epigenetic age estimators.** 'Epigenetic age acceleration' is a term coined to indicate the deviation of chronological age from age as estimated by an 'epigenetic clock' and is associated with disease risk and mortality (47). In our own sample (time-points 6/7 to 17 years), the median absolute difference (MAD) between epigenetic age and chronological age for two often-used epigenetic clocks, those of Horvath (39) ( $n = 353$  CpGs) and Hannum *et al.* (40) ( $n = 71$ ), was rather large but similar to most other published studies (39,40,48,49), with an MAD of 2.9 and 3.8 years, respectively. A previous study found epigenetic age to be non-linearly related to chronological age in both childhood and adulthood (50). However, to use epigenetic age acceleration as an unbiased predictor, one would expect a linear association between the

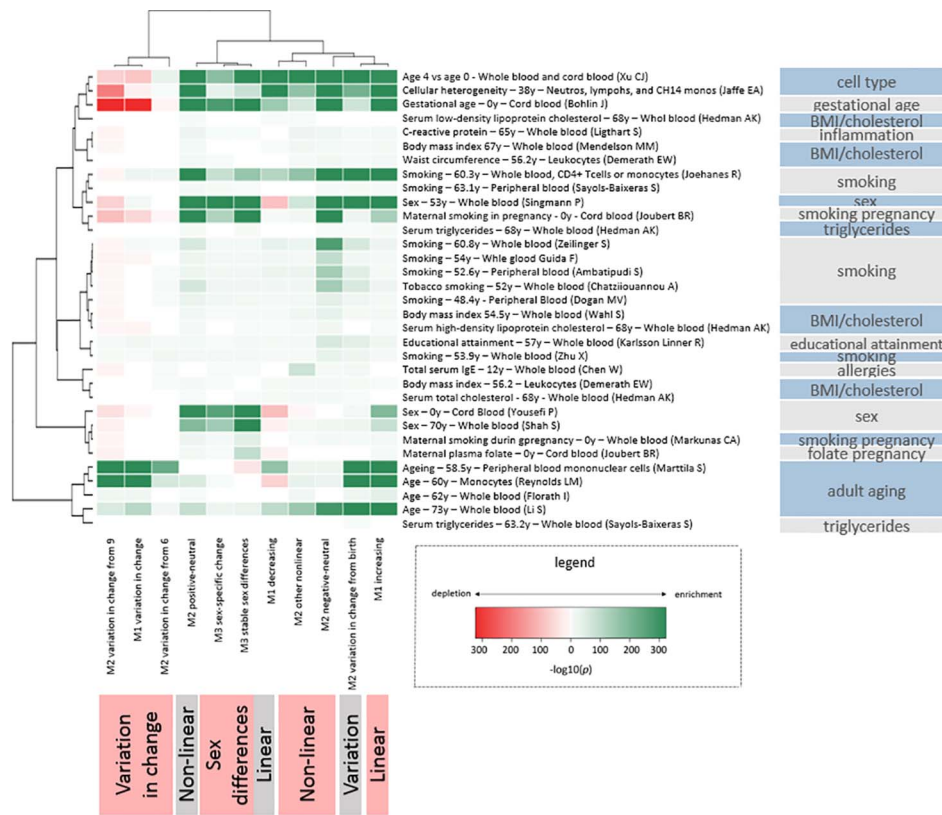


Figure 6. Enrichment of age-related trajectories in EWASs.

epigenetic clock and age, and thus for the clock to contain few CpGs that change nonlinearly over time. Additionally, to detect age acceleration, one would expect that change in DNAm at these CpGs would also vary between individuals. Here, we tested the enrichment of our DNAm trajectories in the epigenetic clocks of Horvath ( $n = 353$  CpGs) and Hannum *et al.* ( $n = 71$ ).

CpGs in both epigenetic clocks more often showed non-linear DNAm patterns than other sites did (26.4% in Horvath's clock showed non-linear patterns; 45.1% in Hannum's clock did). Slope changes in DNAm were found both at the age of 6 (21% in Horvath; 40.8% in Hannum) and 9 years (7.1% in Horvath, 8.4% in Hannum). No enrichment for inter-individual variation in DNAm change was found within Horvath's clock, whereas CpGs in Hannum's clock showed enrichment for inter-individual variation in DNAm change from birth compared with other sites (16.9 versus 3.4%) and from 9 years onwards (28.2 versus 8.2%). Based on this, we conclude neither that both epigenetic clocks did meet our expectations on the specific inclusion of linearly changing probes, nor that CpGs in Horvath's epigenetic clock are enriched for inter-individual variation in DNAm change.

**Functional associations.** Enrichment of GO categories was tested for genes linked to CpGs with different DNAm trajectories. In short, genes annotated to CpGs with overall decreasing DNAm levels were enriched in immune-developmental functions, whereas those annotated to CpGs with increasing levels were enriched in neurodevelopmental functions. This pattern seemed even more pronounced at genes annotated to non-linear Negative-Neutral and Positive-Neutral CpGs, with the former more often associated to immune-development and the latter to neurodevelopment. Genes linked to CpGs with stable sex differences and sex differences in DNAm change were enriched

in pathways associated with sexual development, such as genital development, as well as pathways associated with neurodevelopment. Genes linked to CpGs with sex difference in DNAm change were also enriched in functions related to tooth and hair development.

**Enrichment in EWASs.** We further investigated the functional relevance of CpG sites with age-related DNAm trajectories by testing enrichment with published EWAS associations (Fig. 6) (28,51–76). Unsupervised clustering of the enrichments shows that CpG sites with inter-individual variation in change over time have distinct enrichments and cluster differently from those with age-associated change that is consistent among individuals. The CpG sites of each age-associated DNAm trajectory were enriched with published age associations in adulthood. Multiple smoking EWAS clustered together with enrichment patterns exhibiting strongest enrichments among CpG sites with negative-neutral trajectories and mostly weak enrichments among CpG sites with inter-individual variation in change. Furthermore, despite adjusting for cell count heterogeneity in our models, we observed enrichments of CpG sites that differ by white blood cell (WBC) type among sites following nearly all age-associated trajectories; however, the increases in proportions were small. Finally, we observed enrichments of CpG sites associated with gestational age and prenatal smoking with sex-specific DNAm.

## Discussion

In this study, we described changes in DNAm levels through the first two decades of human life. We examined DNAm levels per CpG by their linear association with age, their non-linear



trajectories and inter-individual variation in change, as well as sex differences and CpG characteristics.

We found that about half of sites change: consistent linear and/or non-linear DNAm change was found at 53% of sites. We further found that over a quarter of sites, 28% were characterized by substantial inter-individual differences in the direction of this change. DNAm sex differences were present, but not abundant: 5% of autosomal sites displayed different DNAm levels or differences in change over time for girls and boys.

Specifically, we determined that DNAm at 52% of the measured methylome have some form of linear change from birth to late adolescence, with DNAm decreasing at 36% and increasing at 16% of CpGs. CpGs with decreasing DNAm tended to have high levels of DNAm and were more often located in gene bodies. CpGs with increasing levels of DNAm tended to have low levels of DNAm and were more likely to be located in promoter regions and at enhancers. The predominance of decreasing CpGs is in agreement with literature on epigenome-wide DNAm and age in cross-sectional research on children and adults (18,77), as well as with longitudinal research in adults (19).

Non-linear DNAm trajectories were detected at 11% of CpGs, mostly involving changes in DNAm from birth to age 6 years, after which DNAm was more stable. We note that this could be due to cord blood being used to generate DNAm profiles at birth, whereas peripheral blood was used at later ages. A previous study (23) including eight children showed that the cord blood DNAm profile at birth clustered separately from later peripheral profiles, after which DNAm changed gradually from 1 to 2.5–5 years. Such differences between DNAm in cord and peripheral blood might be due to uncaptured differences in WBC composition, as well as to different gene-regulatory functioning in the intra-uterine versus extra-uterine environment.

Sites with decreasing levels of DNAm, both with or without slope changes around the age of 6 years, were functionally enriched for immune-developmental pathways, and sites with increasing levels of DNAm, both with or without slope changes, were enriched for neurodevelopmental pathways. Since these observations were based on blood DNAm, it remains to be studied what roles genes linked to neurodevelopmental pathways play in in blood, or to what extent DNAm trajectories in blood mirror those in neural tissue.

To our knowledge, we are the first to study inter-individual differences in DNAm change. Inter-individual differences in linear DNAm trajectories were found at 27% of CpGs, indicating change at different rates or directions for different individuals. Such sites tended to have overall increasing rather than decreasing levels of DNAm from birth to 18 years. More research would be needed to understand why these to qualities so often co-occur in CpGs. Inter-individual differences in non-linear DNAm trajectories were most often found in the slope change at 9 years (8% of CpG sites), indicating that most inter-individual differences in DNAm emerge after the first decade of life. More research is needed to understand if the direction of change in this period is determined by stimuli during that period, or rather by preceding, perhaps cumulative, exposures. However, it is clear that, given the high proportion of CpG sites with inter-individual variation in DNAm change over time that we have observed, it is important to restrict the range of ages of children included a single EWAS. Specific limits should be discussed given the rapidly growing number of studies generating DNAm profiles across childhood (78).

Stable sex differences were found at 5% of autosomal CpGs, and sex differences in DNAm change were found at 0.4% of

all CpGs. In general, if there were stable sex differences, girls had higher levels of DNAm (4% of all CpGs); in case of sex differences in DNAm change, boys had an accelerated upward change (0.2% of all CpGs). The direction of stable sex differences detected is congruent with a cross-sectional study on newborns, in which girls had higher DNAm levels than boys for the large majority of the 3031 significant autosomal CpGs (69). Sex-discordant associations with age seemed to be more prevalent from birth to age 6 years than afterwards, suggesting that any phenotypic sex differences associated to DNAm would be established in early childhood. Sex-associated CpGs showed less inter-individual variation in change and tended to be associated with meQTLs, and therefore, more likely to be genetically influenced. Their enrichment in the shores of CpG islands, areas at which DNAm has been associated with tissue differentiation and tissue-specific gene expression (79), is consistent with the critical role that these processes play in sexual differentiation. Studies into sex differences in epigenetic regulation might want to focus on these locations.

We also found the other DNAm trajectories to be arranged throughout the genome in a non-random fashion. Earlier studies (41,80) have shown that, for active genes, lower DNAm towards the promoter area (TSS200) and higher DNAm in the gene body relate to increased gene transcription. Here, we add the observation that promoter DNAm tends to increase and gene body DNAm tends to decrease with age. From this finding, one might infer that a downregulation of gene expression takes place from birth to late adolescence. Enrichment analyses of published EWAS associations further showed that different traits and exposures exhibited distinct enrichment patterns among DNAm trajectories. For example, there were clear differences between smoking and BMI-related traits. Enrichment of sites with DNAm sex differences in EWASs on prenatal maternal smoking is consistent with studies finding that prenatal smoking affects traits, such as birth weight (81), brain development (82,83) and attention (84) differently in boys and girls. Clustering for prenatal maternal smoking EWASs also showed enrichment for CpGs with consistent change among individuals, not for CpGs with inter-individual variation in change. This may suggest a link with the well-known effects of prenatal smoking on childhood development since consistent DNAm change is more likely related to development or aging programming than inter-individual variation. This may explain why changes associated to prenatal smoking persist throughout life (85). Notably, this pattern of change without inter-individual variation is visible in cg05575921, the AHRR CpG site strongly and persistently associated with prenatal smoking (86,87) (Supplementary Material, Fig. S3; <http://epidelta.mrcieu.ac.uk>).

As a *post hoc* sensitivity analysis, we recomputed our main results, excluding a list of cross-reactive probes (46), and compared these to our main results. In most cases, percentages of CpGs representing the various DNAm trajectories were significantly different when excluding the cross-reactive probes, but absolute percentages did not change more than 0.5%. The trajectory that had a somewhat larger relative change in percentage was the CpGs in which stable sex differences were detected (4.8% in all CpGs; 4.5% excluding cross-reactive probes), which is in line with earlier reports that autosomal cross-reactive probes may reflect false DNAm differences between the sexes (46,88). However, since we found that this group of CpGs is enriched for associated genes involved in sexual development, we conclude that the large majority of the CpGs identified to have sex differences are not affected by cross-reactivity.

We found that DNAm trajectories for CpGs included in epigenetic clocks employed to estimate age or age acceleration did not meet our expectations. For example, we observe that over one-quarter and nearly one-half of the CpG sites included in the Horvath and Hannum clocks, respectively, follow non-linear DNAm trajectories in childhood. Given the widespread use of clocks to investigate biological aging, further investigation is warranted to better understand how, and perhaps if, associations using these clocks should be interpreted in child DNAm profiles.

We note three main limitations of our findings. First, the use of different tissue types (cord blood and peripheral blood) could account for some of the differences between birth and later time points, e.g. sites that increased or decreased between birth and 6, but did not show change after that. Although enrichment analyses further showed that these sites were enriched with sites differentially methylated between WBC types, the enrichments represent small proportions of sites with linear or non-linear change identified in our study. Generation of DNAm profiles of a single tissue type or cell type collected across childhood, or the measurement of WBC proportions instead of estimation of proportions, would be needed to disentangle this issue further. Second, since DNAm at 9 years was measured only in Generation R and at 17 years only in ALSPAC, DNAm differences from 9 to 17 may be to some extent driven by batch effects or cohort differences. This may explain some of the inter-individual differences in slope changes at 9 towards 17 years. However, the high level of agreement in both stability and change among the corresponding time points of the two cohorts is reassuring. Moreover, it is not entirely surprising that inter-individual variation in directionality of change was higher for the largest age interval. This interval, furthermore, encompasses the period of adolescent development, a time in which many inter-individual phenotypic differences arise. Finally, it should be noted that the current study only included children of European ancestry. Considerable DNAm differences have been found between populations (89–91), but research on age-associated DNAm differences is scarce. One study (92) reported evidence for overlap in age-associated CpGs in two African populations with studies on European-ancestry populations, but more research is needed to map the generalizability of longitudinal DNAm changes among different populations.

## Conclusions

In the first comprehensive CpG-by-CpG characterization of DNAm from birth to late adolescence, we found that DNAm at more than half of the studied CpG sites changes consistently between individuals and that considerable inter-individual variation in change exists. Furthermore, characteristics, such as child sex, CpG location, genetic variants, and environmental and disease traits, have distinct associations with patterns of DNAm change. Further analysis of these patterns is made readily available at <http://epidelta.mrcieu.ac.uk/>, which we hope can be used in future studies to test developmental hypotheses that promote our understanding of the developmental nature of DNAm, its role in gene functioning and the associated biological pathways leading to health and disease.

## Materials and Methods

### Setting

Data were obtained from two population-based prospective birth cohorts: the Dutch Generation R Study (Generation R) and

the British Avon Longitudinal Study of Parents and Children (ALSPAC). Pregnant women residing in the study area of Rotterdam, The Netherlands, with an expected delivery date between April 2002 and January 2006 were invited to enrol in Generation R. A more extensive description of the study can be found elsewhere (93). The Generation R Study is conducted in accordance with the World Medical Association Declaration of Helsinki and has been approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam. Informed consent was obtained for all participants.

Pregnant women residing in the study area of former county Avon, UK, with an expected delivery date between April 1991 and December 1992 were invited to enrol in the ALSPAC study. Detailed information on the study design can be found elsewhere (3,94). The ALSPAC website contains the details of all available data through a fully searchable data dictionary and variable search tool (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.

### Study population

In the Generation R Study, 9778 pregnant mothers had 9749 live-born children. For a subsample of 1414 children, DNAm data were collected at birth and/or 6 years and/or 10 years of age. This subsample consisted of participants with parents born in the Netherlands [European ancestry (95) confirmed for all children with genetic data available (95.4%)]. Fifteen sibling pairs were present in the dataset. From each pair, one sibling with the lowest number of DNAm measurements, or otherwise randomly, was excluded, resulting in a sample with 1399 children (with 2333 DNAm samples; see below).

In the ALSPAC study, 15 247 pregnant mothers gave birth to 14 973 live-born children. DNAm at birth and/or 7 years and/or 17 years was available for a subsample of 1003 children as part of the Accessible Resource for Integrated Epigenomic Studies (ARIES) study (96). From this sample, 48 children with non-European ancestry as based on genetic principle component analysis and 6 children with missing data on gestational age were excluded, resulting in a sample of 949 children with DNAm data (with 2686 DNAm samples; see below).

### DNA methylation

Cord blood was drawn after birth for both cohorts, and peripheral blood was drawn at a mean age of 6.0 ( $SD=0.5$ ) and 9.8 ( $SD=0.3$ ) years for Generation R, and 7.5 ( $SD=0.2$ ) and 17.1 ( $SD=1.0$ ) years for ALSPAC. Both cohorts made use of the EZ-96 DNAm kit (shallow) (Zymo Research Corporation, Irvine, CA) to perform bisulfite conversion on the extracted leukocytic DNA. Samples were further processed with the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, CA) to analyze DNAm.

In Generation R, quality control was performed on all 2467 available DNAm samples with the CPACOR workflow (97). Arrays with observed technical problems, such as failed bisulfite conversion, hybridization or extension as well as arrays with a mismatch between sex of the proband and sex determined by the chromosome X and Y probe intensities, were removed

from subsequent analyses. Additionally, only arrays with a call rate >95% per sample were processed further, resulting in 2355 samples, 22 of which belonged to half of an excluded sibling pair; hence, 2333 samples were carried forward into normalization.

In ALSPAC, quality control was performed on 6057 samples (3286 belonging to children, 2771 to their mothers), using the *meffil* package (98) in R version 3.4.3 (99). After removing samples with mismatched genotypes, mismatched sex, incorrect relatedness, low concordance with samples collected at other time points, extreme dye bias and poor probe detection, 5337 samples remained, 2845 of which belonging to children, used in the current study.

Although we removed low-quality samples in part based on high proportions of probes with undetectable signal or low bead numbers, we did not afterward remove any probes due to poor signal. If we apply a typical probe-exclusion procedure, with detection P-value >0.01 or low bead count for more than 10% of the samples, we would remove only 1414 or 0.3% of all autosomal probes. The removal of such a small number of probes would not change our results in any noteworthy way.

To minimize cohort effects as much as possible, we normalized both cohorts together as a single dataset. Functional normalization (10 control probe principal components, slide included as a random effect) was performed with the *meffil* package in R (98). Normalization took place on the combined Generation R and ALSPAC set comprising a total of 5178 samples for a total of 485 512 CpGs. One-hundred and fifty-nine ALSPAC samples belonging to non-European children or children with missing data on gestational age were excluded, leading to a final ALSPAC set of 2686 samples (for 949 children). Together with 2333 samples for Generation R (of 1399 children), they formed a combined set of 5019 samples (of 2348 children.)

Analyses were restricted to 473 864 autosomal CpGs. DNAm levels were operationalized as beta values ( $\beta$  values), representing the ratio of methylated signal relative to the sum of methylated and unmethylated signal measured per CpG. In a post hoc sensitivity analysis, recomputed our main results excluding a list of cross-reactive probes [ $n=28\,298$ ; 5.9% of all included CpGs (46)] and compared them to our main results (Supplementary Table S16).

## Covariates

Sample plate number ( $N=29$  in Generation R and  $N=31$  in ALSPAC) was used to correct for batch effects, which was added as a random variable in the model (see below). WBC composition was estimated with the reference-based Bakulski method (100) for cord blood and Houseman method (101) for peripheral blood (Supplementary Material, Table S17). Nucleated red blood cells were not further analyzed due to its specificity to cord blood, leaving CD4+ T-lymphocytes, CD8+ T-lymphocytes, natural killer cells, B-lymphocytes, monocytes and granulocytes. Other covariates included gestational age in weeks, sex of the child, and cohort.

## Statistical analyses

*Step 1: Assessing cross-cohort comparability in DNA methylation stability.*

To ascertain comparability among the two cohorts, we compared within-cohort DNAm stability between the time points that were present in both cohorts—i.e. birth and 6/7 years (Generation R/ALSPAC, respectively).

Longitudinal stability per CpG within each cohort was assessed by studying estimates of concordance and change.

For concordance, DNAm data were first residualized within each cohort for all variables present in the longitudinal models except the 'cohort' variable, in order to remove between-cohort differences due to other covariates. Concordance was then measured both with Spearman correlation (data at most CpGs is not normally distributed) as a measure of relative concordance and with intra-class correlations as a measure of absolute concordance (children with data for both time points:  $n$  Generation R = 476,  $n$  ALSPAC = 826). Longitudinal change from birth to 6/7 years was assessed by studying the estimates of the change in DNAm per year by applying Model 1 (see below) within each cohort (children with data for at least one of the two time-points:  $n$  Generation R = 1394,  $n$  ALSPAC = 944).

In a second step, cross-cohort comparability was assessed with Spearman ( $\rho$ ) correlation of concordance estimates of the CpGs of each cohort (which were not normally distributed) and Pearson correlations ( $r$ ) among the change estimates of the CpGs of each cohort (which were normally distributed).

*Step 2: Longitudinal modelling of DNA methylation using combined Generation R and ALSPAC data.*

The combined Generation R and ALSPAC dataset had four time points of collection (birth, age 6/7, 10 and 17 years). We fit three linear mixed models to CpG site DNAm across the genome to identify (i) linear change over time (Model 1); (ii) non-linear change over time (Model 2) and (iii) sex differences in change over time (Model 3). Both fixed and random effects were examined to allow for inter-individual variation in DNAm patterns over time. The models are described in detail below.

*Model 1: Linear change.* This model was applied to identify CpGs that show an overall change in DNAm from birth to 18 years (i.e. fixed age effect), as well as CpGs with inter-individual differences in change during that time (i.e. random age effect). The Model 1 is defined as follows:

$$\begin{aligned} M1: M_{ijk} &= \beta_0 + u_{0i} + \beta_1 \text{Age}_{ij} + u_{1i} \text{Age}_{ij} + u_{0k} + \text{covariates} + \epsilon_{ijk} \\ \epsilon_{ijk} &\sim N(0, \sigma_{\epsilon}^2) \\ u_{0i} &\sim N(0, \sigma_{0i}^2) \\ u_{1i} &\sim N(0, \sigma_{1i}^2) \\ u_{0k} &\sim N(0, \sigma_{0k}^2) \end{aligned}$$

(i.e.:  $\text{CpG} \sim (1|\text{Child ID}) + \text{Age} + (\text{Age}|\text{Child ID}) + (1|\text{Sample Plate}) + \text{covariates}$ ).

Here, participants are denoted by  $i$ , time points by  $j$  and sample plates by  $k$ .  $M$  denotes DNAm level,  $\beta_0$  fixed intercept,  $u_{0i}$  random intercept,  $\beta_1$  fixed age coefficient,  $u_{1i}$  random age coefficient and  $u_{0k}$  random intercept for sample plate. Hence,  $\beta_1$  represents the average change in DNAm per one year. Variability in this change among individuals was captured with  $u_{1i}$ . To avoid problems with model identification, the random slope of age was uncorrelated to the random intercept (i.e. a diagonal random effects matrix was used).

*Model 2: Non-linear change.* To identify non-linear changes in DNAm, we extended Model 1 to allow slope changes at ages 6 and 9 (30,31)

$$\begin{aligned} M2: M_{ijk} &= \beta_0 + u_{0i} + \beta_1 \text{Age}_{ij} + \beta_2 (\text{Age}_{ij}-6)^+ + \beta_3 (\text{Age}_{ij}-9)^+ + u_{1i} \\ &\text{Age}_{ij} + u_{2i} (\text{Age}_{ij}-6)^+ + u_{3i} (\text{Age}_{ij}-9)^+ + u_{0k} + \text{covariates} + \epsilon_{ijk} \\ \epsilon_{ijk} &\sim N(0, \sigma_{\epsilon}^2) \\ u_{0i} &\sim N(0, \sigma_{0i}^2) \\ u_{1i} &\sim N(0, \sigma_{1i}^2) \\ u_{2i} &\sim N(0, \sigma_{2i}^2) \\ u_{3i} &\sim N(0, \sigma_{3i}^2) \\ u_{0k} &\sim N(0, \sigma_{0k}^2) \end{aligned}$$

(i.e.  $\text{CpG} \sim (1|\text{Child ID}) + \text{Age} + \text{Age}-6 + \text{Age}-9 + (\text{Age}|\text{Child ID}) + (\text{Age}-6|\text{Child ID}) + (\text{Age}-9|\text{Child ID}) + (1|\text{Sample Plate}) + \text{covariates}$ ),

Where  $a^+ = a$  if  $a > 0$  and 0 otherwise, so that  $\beta_2$  represents the average change in DNAm per year from 6 years of age onward, after accounting for the change per year from birth onward, as denoted by  $\beta_1$ . Likewise,  $\beta_3$  represents the average change in DNAm per year from 9 years of age onward, after accounting for the change per year from 6 years of age onward. Hence, with those variables, we are able to detect slope changes at 6 and 9 years old. These slope changes were used to identify different types of non-linear patterns. With  $u_{2i}$  and  $u_{3i}$ , the inter-individual variation in slope changes at 6 and 9 years were captured, respectively. General linear hypothesis testing (102) was applied to our fitted models to determine if there were changes in DNAm per year from 6 to 9 years and from 9 to 18 years.

**Model 3: Sex differences in change:** To identify CpGs for which DNAm changes differently over time for boys and girls, we applied the following model:

M3:  $M_{ijk} = \beta_0 + u_{0i} + \beta_1 \text{Age}_{ij} + u_{1i} \text{Age}_{ij} + \beta_2 \text{Sex}_i \text{Age}_{ij} + u_{0k} + \text{covariates} + \epsilon_{ijk}$

$$\epsilon_{ijk} \sim N(0, \sigma_{\epsilon}^2)$$

$$u_{0i} \sim N(0, \sigma_{u_0}^2)$$

$$u_{1i} \sim N(0, \sigma_{u_1}^2)$$

$$u_{0k} \sim N(0, \sigma_{u_0k}^2)$$

(i.e.  $\text{CpG} \sim (1|\text{Child ID}) + \text{Age} + (\text{Age}|\text{Child ID}) + \text{Age} * \text{Sex} + (1|\text{Sample Plate}) + \text{covariates}$ ),

Here,  $\text{Sex}_i$  denotes the sex of child  $i$ . Both main and interaction effects for sex were studied.

The three mixed models were fitted using maximum likelihood estimation in R with the *lme4* package (103). Continuous covariates (WBCs, gestational age) were z-score standardized. Random slopes were kept uncorrelated with random intercepts and the NLOpt optimizer was used, enabling us to improve computational speed compared with the default settings. P-values for the fixed effects were computed with a z-test. P-values for random slopes of the Age effects were obtained by refitting the model without the random slope and comparing the fit estimates of the two models with a likelihood ratio test. Within each model, P-value thresholds were Bonferroni-corrected for the number of tested CpGs (i.e. to  $P < 1 \times 10^{-7}$ ).

**Step 3: Functional characterization of CpGs with comparable patterns of change.**

To interpret the functionality of the age-related DNAm patterns from the three models, CpG sites adhering to 8 different age-related patterns (M1 linear change and inter-individual variation in linear change, M2 non-linear trajectories and inter-individual variation in change from birth in slope change at 6 and 9 years, and M3 stable sex differences and sex differences in DNAm change) were tested for enrichment in the following:

- (i) gene-relative genomic regions (TSS1500, TSS200, 5'UTR, 1st exon, gene body, 3'UTR and intergenic regions) as indicated by the UCSC Genome Browser (104) in the Illumina HumanMethylation450 v1.2 Manifest (Illumina Inc.),
- (ii) CpG island-relative genomic regions (N shelf, N shore, CpG island, S shore, S shelf and open sea regions) as indicated by the UCSC Genome Browser (104) in the Illumina HumanMethylation450 v1.2 Manifest (Illumina Inc.),
- (iii) enhancer elements as those expressed in whole blood, peripheral blood mononuclear cells, natural killer cells, CD4+ T cells, CD8+ T cells, monocytes, neutrophils, eosinophils or B cells (105),
- (iv) cis and trans meQTLs identified by the BIOS consortium (45) and polymorphic CpGs (46) and

- (v) inclusion in two well-known epigenetic age estimators (39,40).

Altogether, these encompassed 19 enrichment analyses for eight variables. Enrichment was tested using  $\chi^2$ -tests of unequal proportions. The enrichment P-value threshold was Bonferroni-corrected for multiple tests (i.e.  $P < 3.29 \times 10^{-4}$  for  $8 \times 19 = 152$  tests). Second, we tested the enrichment of GO categories for genes linked to CpG sites surviving adjustment for multiple tests ( $P < 1 \times 10^{-7}$ ) for each of the main variables of interest. Gene annotation (closest genes, <1500 base pairs from transcription start site) was based on the Illumina HumanMethylation450 v1.2 Manifest (Illumina Inc.). The analysis was adjusted for gene size and pruned for near-identical terms [see elsewhere for a full description (106)]. For completeness, terms with nominal  $P < 0.05$  were reported. At last, we tested the enrichment of age-related DNAm trajectories (11 different age-related patterns: M1 decreasing, increasing and inter-individual variation in linear change, M2 Positive-Neutral, Negative-Neutral, other non-linear, inter-individual variation in change from birth, in slope change at 6 and 9 years and M3 stable sex differences and sex differences in DNAm change) in EWASs on age, prenatal smoking, smoking, cardiovascular-associated traits, C-reactive protein, allergies, educational attainment and cellular heterogeneity. EWAS summary statistics were retrieved from the EWAS Catalog (<http://www.ewascatalog.org/>) and studies were included when performed with the 450 K array in peripheral or cord blood, resulting in 33 EWASs. Enrichment was tested with Fisher's exact tests; the enrichment P-value threshold was Bonferroni-corrected for multiple tests (i.e.  $P < 1.38 \times 10^{-4}$  for  $11 \times 33 = 363$  tests).

## Supplementary Material

Supplementary Material is available at HMG online.

## Acknowledgements

The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with the Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC), Rotterdam. We gratefully acknowledge the contribution of children and parents, general practitioners, hospitals, midwives and pharmacies in Rotterdam. The generation and management of the Illumina 450 K methylation array data (EWAS data) for the Generation R Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, The Netherlands. We thank Mr Michael Verbiest, Ms Mila Jhamai, Ms Sarah Higgins, Mr Marijn Verkerk and Dr Lisette Stolk for their help in creating the EWAS database. We thank Dr A. Teumer for his work on the quality control and normalization scripts.

We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

**Conflict of Interest statement.** The authors declare that they have no competing interests.

## Funding

The general design of the Generation R Study is made possible by financial support from the Erasmus Medical Center, Rotterdam, the Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development and the Ministry of Health, Welfare and Sport. The EWAS data were funded by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA; project no. 050-060-810), by funds from the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by a grant from the National Institute of Child and Human Development (R01HD068437). This project further received funding from the European Union's Horizon 2020 research and innovation programme (733206, LifeCycle; 848158, EarlyCause; 874739, LongITools) and from the European Joint Programming Initiative 'A Healthy Diet for a Healthy Life' (JPI HDHL, NutriPROGRAM project, ZonMw the Netherlands no. 529051022 and PREcisE project ZonMw the Netherlands no. 529051023).

The UK Medical Research Council and Wellcome (grant number 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). This publication is the work of the authors and Dr Suderman will serve as a guarantor for the ALSPAC-related contents of this paper.

Analysis of the ALSPAC data was funded by the UK Economic and Social Research Council grant (grant number ES/N000498/1). ARIES was funded by the BBSRC (BBI025751/1 and BB/I025263/1). Supplementary funding to generate DNA methylation data which are (or will be) included in ARIES has been obtained from the MRC, ESRC, NIH and other sources. ARIES is maintained under the auspices of the MRC Integrative Epidemiology Unit at the University of Bristol (grant numbers MC\_UU\_00011/4 and MC\_UU\_00011/5).

A.N. and H.T. are supported by a grant of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant no. 024.001.003, Consortium on Individual Development). H.T. was also supported by a grant from the Netherlands Organization for Scientific Research (NWO-grant 016.VICI.170.200). E.W. is funded by CLOSER, who is funded by ESRC and MRC (grant reference: ES/K000357/1). The funders took no role in the design, execution, analysis or interpretation of the data or in the writing up of the findings. [www.closer.ac.uk](http://www.closer.ac.uk). The work of CAMC has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 707404. L.C.H. is supported by ESRC/BBSRC Grant ES/N000382/1 for the Interpreting epigenetic signatures in studies of early life adversity project (Interstela Project). J.R. is supported by the Netherlands Organization for Scientific Research (NWO ZonMw VENI, grant no. 91618147). V.W.V.J. received funding from the European Research Council (ERC-2014-CoG-648916). M.J.B.-K. is supported by the European Research Council (ERC AdG) and by the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant No. 024.001.003, Consortium on Individual Development). M.H.v.I.J. is supported by the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant No. 024.001.003, Consortium on Individual Development) and by a Spinoza Award of the Netherlands Organization for Scientific Research.

## References

1. Khavari, D.A., Sen, G.L. and Rinn, J.L. (2010) DNA methylation and epigenetic control of cellular differentiation. *Cell Cycle*, **9**, 3880–3883.
2. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C. and Jaffe, D.B. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766–770.
3. Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L. and Ness, A. (2012) Cohort profile: the Avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.*, **42**, 97–110.
4. Chow, J. and Heard, E. (2009) X inactivation and the complexities of silencing a sex chromosome. *Curr. Opin. Cell Biol.*, **21**, 359–366.
5. Koukoura, O., Sifakis, S. and Spandidos, D.A. (2012) DNA methylation in the human placenta and fetal growth. *Mol. Med. Rep.*, **5**, 883–889.
6. Eden, A., Gaudet, F., Waghmare, A. and Jaenisch, R. (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. *Science*, **300**, 455–455.
7. Comet, I., Riising, E.M., Leblanc, B. and Helin, K. (2016) Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat. Rev. Cancer*, **16**, 803.
8. Natoli, G. (2010) Maintaining cell identity through global control of genomic organization. *Immunity*, **33**, 12–24.
9. Bojesen, S.E., Timpson, N., Relton, C., Smith, G.D. and Nordestgaard, B.G. (2017) AHRH (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax*, **72**, 646–653.
10. Cao-Lei, L., Massart, R., Suderman, M.J., Machnes, Z., Elgbeili, G., Laplante, D.P., Szyf, M. and King, S. (2014) DNA methylation signatures triggered by prenatal maternal stress exposure to a natural disaster: project ice storm. *PLoS One*, **9**, e107653.
11. Baccarelli, A., Wright, R.O., Bollati, V., Tarantini, L., Litonjua, A.A., Suh, H.H., Zanobetti, A., Sparrow, D., Vokonas, P.S. and Schwartz, J. (2009) Rapid DNA methylation changes after exposure to traffic particles. *Am. J. Resp. Crit. Care*, **179**, 572–578.
12. Suarez-Alvarez, B., Rodriguez, R.M., Fraga, M.F. and López-Larrea, C. (2012) DNA methylation: a promising landscape for immune system-related diseases. *Trends Genet.*, **28**, 506–514.
13. Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M. and Levey, A.I. (2011) 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.*, **14**, 1607.
14. Gaunt, T.R., Shihab, H.A., Hemani, G., Min, J.L., Woodward, G., Lyttleton, O., Zheng, J., Duggirala, A., McArdle, W.L. and Ho, K. (2016) Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.*, **17**, 1–14.
15. Feil, R. and Fraga, M.F. (2012) Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.*, **13**, 97.
16. Bell, J.T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F. and Valdes, A. (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.*, **8**, e1002629.

17. 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. and Traynor, B.J. (2011) Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum. Mol. Genet.*, **20**, 1164–1172.
18. Alisch, R.S., Barwick, B.G., Chopra, P., Myrick, L.K., Satten, G.A., Conneely, K.N. and Warren, S.T. (2012) Age-associated DNA methylation in pediatric populations. *Genome Res.*, **22**, 623–632.
19. Wang, Y., Karlsson, R., Lampa, E., Zhang, Q., Hedman, Å.K., Almgren, M., Almqvist, C., McRae, A.F., Marioni, R.E. and Ingelsson, E. (2018) Epigenetic influences on aging: a longitudinal genome-wide methylation study in old Swedish twins. *Epigenetics*, **13**, 975–987.
20. Wang, Y., Pedersen, N.L. and Hägg, S. (2018) Implementing a method for studying longitudinal DNA methylation variability in association with age. *Epigenetics*, **13**, 866–874.
21. Lévesque, M.L., Casey, K.F., Szyf, M., Ismaylova, E., Ly, V., Verner, M.-P., Suderman, M., Brendgen, M., Vitaro, F. and Dionne, G. (2014) Genome-wide DNA methylation variability in adolescent monozygotic twins followed since birth. *Epigenetics*, **9**, 1410–1422.
22. Martino, D.J., Loke, Y.J., Gordon, L., Ollikainen, M., Cruickshank, M.N., Saffery, R. and Craig, J.M. (2013) 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.*, **14**, R42.
23. Martino, D.J., Tulic, M.K., Gordon, L., Hodder, M., Richman, T.R., Metcalfe, J., Prescott, S.L. and Saffery, R. (2011) Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics*, **6**, 1085–1094.
24. Fitzmaurice, G.M., Laird, N.M. and Ware, J.H. (2012) *Applied Longitudinal Analysis*. John Wiley & Sons, Hoboken, New Jersey.
25. Ng, J.W.Y., Barrett, L.M., Wong, A., Kuh, D., Smith, G.D. and Relton, C.L. (2012) The role of longitudinal cohort studies in epigenetic epidemiology: challenges and opportunities. *Genome Biol.*, **13**, 246.
26. Naumova, E.N., Must, A. and Laird, N.M. (2001) Tutorial in biostatistics: evaluating the impact of ‘critical periods’ in longitudinal studies of growth using piecewise mixed effects models. *Int. J. Epidemiol.*, **30**, 1332–1341.
27. Cudeck, R. and Harring, J.R. (2007) Analysis of nonlinear patterns of change with random coefficient models. *Annu. Rev. Psychol.*, **58**, 615–637.
28. Xu, C.-J., Bonder, M.J., Söderhäll, C., Bustamante, M., Baiz, N., Gehring, U., Jankipersadsing, S.A., Van Der Vlies, P., van Diemen, C.C. and Van Rijkom, B. (2017) The emerging landscape of dynamic DNA methylation in early childhood. *BMC Genomics*, **18**, 1–11.
29. Richmond, R.C., Simpkin, A.J., Woodward, G., Gaunt, T.R., Lyttleton, O., McArdle, W.L., Ring, S.M., Smith, A.D.A.C., Timpson, N.J. and Tilling, K. (2014) Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Hum. Mol. Genet.*, **24**, 2201–2217.
30. Simpkin, A.J., Suderman, M., Gaunt, T.R., Lyttleton, O., McArdle, W.L., Ring, S.M., Tilling, K., Davey Smith, G. and Relton, C.L. (2015) Longitudinal analysis of DNA methylation associated with birth weight and gestational age. *Hum. Mol. Genet.*, **24**, 3752–3763.
31. Sharp, G.C., Lawlor, D.A., Richmond, R.C., Fraser, A., Simpkin, A., Suderman, M., Shihab, H.A., Lyttleton, O., McArdle, W. and Ring, S.M. (2015) Maternal pre-pregnancy BMI and gestational weight gain, offspring DNA methylation and later offspring adiposity: findings from the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.*, **44**, 1288–1304.
32. Blatch, G.L. and Lässle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*, **21**, 932–939.
33. Gronostajski, R.M. (2000) Roles of the NFI/CTF gene family in transcription and development. *Gene*, **249**, 31–45.
34. Sheng, M. and Kim, E. (2000) The shank family of scaffold proteins. *J. Cell Sci.*, **113**, 1851–1856.
35. Flick, M.J. and Konieczny, S.F. (2000) The muscle regulatory and structural protein MLP is a cytoskeletal binding partner of beta-spectrin. *J. Cell Sci.*, **113**, 1553–1564.
36. Zaslavsky, K., Zhang, W.-B., McCready, F.P., Rodrigues, D.C., Deneault, E., Loo, C., Zhao, M., Ross, P.J., El Hajjar, J. and Romm, A. (2019) SHANK2 mutations associated with autism spectrum disorder cause hyperconnectivity of human neurons. *Nat. Neurosci.*, **22**, 556–564.
37. Geier, C., Gehmlich, K., Ehler, E., Hassfeld, S., Perrot, A., Hayess, K., Cardim, N., Wenzel, K., Erdmann, B. and Krackhardt, F. (2008) Beyond the sarcomere: CSRP3 mutations cause hypertrophic cardiomyopathy. *Hum. Mol. Genet.*, **17**, 2753–2765.
38. Klaassens, M., Morrogh, D., Rosser, E.M., Jaffer, F., Vreeburg, M., Bok, L.A., Segboer, T., Van Belzen, M., Quinlivan, R.M. and Kumar, A. (2015) Malan syndrome: sotos-like overgrowth with de novo NF1X sequence variants and deletions in six new patients and a review of the literature. *Eur. J. Hum. Genet.*, **23**, 610–615.
39. Horvath, S. (2013) DNA methylation age of human tissues and cell types. *Genome Biol.*, **14**, 3156.
40. Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sada, S., Klotzle, B., Bibikova, M., Fan, J.-B. and Gao, Y. (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell*, **49**, 359–367.
41. Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E.M., Park, I.-H., Xie, B., Daley, G.Q. and Church, G.M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.*, **27**, 361.
42. Wagner, J.R., Busche, S., Ge, B., Kwan, T., Pastinen, T. and Blanchette, M. (2014) The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol.*, **15**, R37.
43. Illingworth, R.S. and Bird, A.P. (2009) CpG islands—‘a rough guide’. *FEBS Lett.*, **583**, 1713–1720.
44. Ong, C.-T. and Corces, V.G. (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.*, **12**, 283–293.
45. Bonder, M.J., Luijk, R., Zhernakova, D.V., Moed, M., Deelen, P., Vermaat, M., Van Ijzerman, M., Van Dijk, F., Van Galen, M. and Bot, J. (2017) Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.*, **49**, 131–138.
46. Chen, Y.-A., Lemire, M., Choufani, S., Butcher, D.T., Grafo-datskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J. and Weksberg, R. (2013) Discovery of cross-reactive probes and

- polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*, **8**, 203–209.
47. Horvath, S. and Raj, K. (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.*, **19**, 371.
  48. McEwen, L.M., Jones, M.J., Lin, D.T.S., Edgar, R.D., Husquin, L.T., MacIsaac, J.L., Ramadori, K.E., Morin, A.M., Rider, C.F. and Carlsten, C. (2018) Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium MethylationEPIC BeadChip array. *Clin. Epigenetics*, **10**, 1–9.
  49. Armstrong, N.J., Mather, K.A., Thalamuthu, A., Wright, M.J., Trollor, J.N., Ames, D., Brodaty, H., Schofield, P.R., Sachdev, P.S. and Kwok, J.B. (2017) Aging, exceptional longevity and comparisons of the Hannum and Horvath epigenetic clocks. *Epigenomics*, **9**, 689–700.
  50. Snir, S., Farrell, C. and Pellegrini, M. (2019) Human epigenetic ageing is logarithmic with time across the entire lifespan. *Epigenetics*, **14**, 912–926.
  51. Jaffe, A.E. and Irizarry, R.A. (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.*, **15**, R31.
  52. Bohlin, J., Häberg, S.E., Magnus, P., Reese, S.E., Gjessing, H.K., Magnus, M.C., Parr, C.L., Page, C.M., London, S.J. and Nystad, W. (2016) Prediction of gestational age based on genome-wide differentially methylated regions. *Genome Biol.*, **17**, 207.
  53. Hedman, Å.K., Mendelson, M.M., Marioni, R.E., Gustafsson, S., Joehanes, R., Irvin, M.R., Zhi, D., Sandling, J.K., Yao, C. and Liu, C. (2017) Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results from genome-wide association studies. *Circ. Cardiovasc. Genet.*, **10**, e001487.
  54. Ligthart, S., Marzi, C., Aslibekyan, S., Mendelson, M.M., Conneely, K.N., Tanaka, T., Colicino, E., Waite, L.L., Joehanes, R. and Guan, W. (2016) DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.*, **17**, 255.
  55. Mendelson, M.M., Marioni, R.E., Joehanes, R., Liu, C., Hedman, Å.K., Aslibekyan, S., Demerath, E.W., Guan, W., Zhi, D. and Yao, C. (2017) Association of body mass index with DNA methylation and gene expression in blood cells and relations to cardiometabolic disease: a Mendelian randomization approach. *PLoS Med.*, **14**, e1002215.
  56. Demerath, E.W., Guan, W., Grove, M.L., Aslibekyan, S., Mendelson, M., Zhou, Y.-H., Hedman, Å.K., Sandling, J.K., Li, L.-A. and Irvin, M.R. (2015) Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum. Mol. Genet.*, **24**, 4464–4479.
  57. Joehanes, R., Just, A.C., Marioni, R.E., Pilling, L.C., Reynolds, L.M., Mandaviya, P.R., Guan, W., Xu, T., Elks, C.E. and Aslibekyan, S. (2016) Epigenetic signatures of cigarette smoking. *Circ. Cardiovasc. Genet.*, **9**, 436–447.
  58. Sayols-Baixeras, S., Lluís-Ganella, C., Subirana, I., Salas, L.A., Vilahur, N., Corella, D., Munoz, D., Segura, A., Jimenez-Conde, J. and Moran, S. (2015) Identification of a new locus and validation of previously reported loci showing differential methylation associated with smoking. The REGICOR study. *Epigenetics*, **10**, 1156–1165.
  59. Sayols-Baixeras, S., Subirana, I., Lluís-Ganella, C., Civeira, F., Roquer, J., Do, A.N., Absher, D., Cenarro, A., Muñoz, D. and Soriano-Tárraga, C. (2016) Identification and validation of seven new loci showing differential DNA methylation related to serum lipid profile: an epigenome-wide approach. The REGICOR study. *Hum. Mol. Genet.*, **25**, 4556–4565.
  60. Singmann, P., Shem-Tov, D., Wahl, S., Grallert, H., Fiorito, G., Shin, S.-Y., Schramm, K., Wolf, P., Kunze, S. and Baran, Y. (2015) Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenet. Chromatin*, **8**, 1–13.
  61. Joubert, B.R., Felix, J.F., Yousefi, P., Bakulski, K.M., Just, A.C., Breton, C., Reese, S.E., Markunas, C.A., Richmond, R.C. and Xu, C.-J. (2016) DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. *Am. J. Hum. Genet.*, **98**, 680–696.
  62. Zeilinger, S., Kühnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J. and Peters, A. (2013) Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*, **8**, e63812.
  63. Guida, F., Sandanger, T.M., Castagné, R., Campanella, G., Polidoro, S., Palli, D., Krogh, V., Tumino, R., Sacerdote, C. and Panico, S. (2015) Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum. Mol. Genet.*, **24**, 2349–2359.
  64. Ambatipudi, S., Cuenin, C., Hernandez-Vargas, H., Ghanatous, A., Le Calvez-Kelm, F., Kaaks, R., Barrdahl, M., Boeing, H., Aleksandrova, K. and Trichopoulou, A. (2016) Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics*, **8**, 599–618.
  65. Chatziioannou, A., Georgiadis, P., Hebels, D.G., Liampa, I., Valavanis, I., Bergdahl, I.A., Johansson, A., Palli, D., Chadeau-Hyam, M. and Siskos, A.P. (2017) Blood-based omic profiling supports female susceptibility to tobacco smoke-induced cardiovascular diseases. *Sci. Rep.*, **7**, 42870.
  66. Linnér, R.K., Marioni, R.E., Rietveld, C.A., Simpkin, A.J., Davies, N.M., Watanabe, K., Armstrong, N.J., Auro, K., Baumbach, C. and Bonder, M.J. (2017) An epigenome-wide association study meta-analysis of educational attainment. *Mol. Australas. Psychiatry*, **22**, 1680–1690.
  67. Zhu, X., Li, J., Deng, S., Yu, K., Liu, X., Deng, Q., Sun, H., Zhang, X., He, M. and Guo, H. (2016) Genome-wide analysis of DNA methylation and cigarette smoking in a Chinese population. *Environ. Health Perspect.*, **124**, 966–973.
  68. Chen, W., Wang, T., Pino-Yanes, M., Forno, E., Liang, L., Yan, Q., Hu, D., Weeks, D.E., Baccarelli, A. and Acosta-Perez, E. (2017) An epigenome-wide association study of total serum IgE in Hispanic children. *J. Allergy Clin. Immunol.*, **140**, 571–577.
  69. Yousefi, P., Huen, K., Davé, V., Barcellos, L., Eskenazi, B. and Holland, N. (2015) Sex differences in DNA methylation assessed by 450 K BeadChip in newborns. *BMC Genomics*, **16**, 911.
  70. Shah, S., McRae, A.F., Marioni, R.E., Harris, S.E., Gibson, J., Henders, A.K., Redmond, P., Cox, S.R., Pattie, A. and Corley, J. (2014) Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.*, **24**, 1725–1733.
  71. Markunas, C.A., Xu, Z., Harlid, S., Wade, P.A., Lie, R.T., Taylor, J.A. and Wilcox, A.J. (2014) Identification of DNA methylation changes in newborns related to maternal smoking during pregnancy. *Environ. Health Perspect.*, **122**, 1147–1153.
  72. Joubert, B.R., Herman, T., Felix, J.F., Bohlin, J., Ligthart, S., Beckett, E., Tiemeier, H., Van Meurs, J.B., Uitterlinden, A.G. and Hofman, A. (2016) Maternal plasma folate impacts

- differential DNA methylation in an epigenome-wide meta-analysis of newborns. *Nat. Commun.*, **7**, 10577.
73. Marttila, S., Kananen, L., Häyrynen, S., Jylhävä, J., Nevalainen, T., Hervonen, A., Jylhä, M., Nykter, M. and Hurme, M. (2015) Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics*, **16**, 179.
  74. Reynolds, L.M., Taylor, J.R., Ding, J., Lohman, K., Johnson, C., Siscovick, D., Burke, G., Post, W., Shea, S. and Jacobs, D.R., Jr. (2014) Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat. Commun.*, **5**, 1–8.
  75. Florath, I., Butterbach, K., Müller, H., Bewerunge-Hudler, M. and Brenner, H. (2014) Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum. Mol. Genet.*, **23**, 1186–1201.
  76. Li, S., Christiansen, L., Christensen, K., Kruse, T.A., Redmond, P., Marioni, R.E., Deary, I.J. and Tan, Q. (2017) Identification, replication and characterization of epigenetic remodelling in the aging genome: a cross population analysis. *Sci. Rep.*, **7**, 1–8.
  77. Johansson, Å., Enroth, S. and Gyllensten, U. (2013) Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS One*, **8**, e67378.
  78. Felix, J.F., Joubert, B.R., Baccarelli, A.A., Sharp, G.C., Almqvist, C., Annesi-Maesano, I., Arshad, H., Baiz, N., Bakermans-Kranenburg, M.J. and Bakulski, K.M. (2017) Cohort profile: pregnancy and childhood epigenetics (PACE) consortium. *Int. J. Epidemiol.*, **47**, 22–23u.
  79. Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M. and Webster, M. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.*, **41**, 178–186.
  80. Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, **13**, 484–492.
  81. Voigt, M., Hermanussen, M., Wittwer-Backofen, U., Fusch, C. and Hesse, V. (2006) Sex-specific differences in birth weight due to maternal smoking during pregnancy. *Eur. J. Pediatr.*, **165**, 757–761.
  82. Paus, T., Nawazkhan, I., Leonard, G., Perron, M., Pike, G.B., Pitiot, A., Richer, L., Veillette, S. and Pausova, Z. (2008) Corpus callosum in adolescent offspring exposed prenatally to maternal cigarette smoking. *Neuroimage*, **40**, 435–441.
  83. Toro, R., Leonard, G., Lerner, J.V., Lerner, R.M., Perron, M., Pike, G.B., Richer, L., Veillette, S., Pausova, Z. and Paus, T. (2008) Prenatal exposure to maternal cigarette smoking and the adolescent cerebral cortex. *Neuropsychopharmacology*, **33**, 1019–1027.
  84. Jacobsen, L.K., Slotkin, T.A., Mencl, W.E., Frost, S.J. and Pugh, K.R. (2007) Gender-specific effects of prenatal and adolescent exposure to tobacco smoke on auditory and visual attention. *Neuropsychopharmacology*, **32**, 2453–2464.
  85. Richmond, R.C., Suderman, M., Langdon, R., Relton, C.L. and Davey Smith, G. (2018) DNA methylation as a marker for prenatal smoke exposure in adults. *Int. J. Epidemiol.*, **47**, 1120–1130.
  86. Kupers, L.K., Xu, X., Jankipersadsing, S.A., Vaez, A., la Bastide-van Gemert, S., Scholtens, S., Nolte, I.M., Richmond, R.C., Relton, C.L., Felix, J.F. et al. (2015) DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int. J. Epidemiol.*, **44**, 1224–1237.
  87. Nielsen, C.H., Larsen, A. and Nielsen, A.L. (2016) DNA methylation alterations in response to prenatal exposure of maternal cigarette smoking: a persistent epigenetic impact on health from maternal lifestyle? *Arch. Toxicol.*, **90**, 231–245.
  88. Chen, Y.-A., Choufani, S., Grafodatskaya, D., Butcher, D.T., Ferreira, J.C. and Weksberg, R. (2012) Cross-reactive DNA microarray probes lead to false discovery of autosomal sex-associated DNA methylation. *Am. J. Hum. Genet.*, **91**, 762–764.
  89. Heyn, H., Moran, S., Hernando-Herraez, I., Sayols, S., Gomez, A., Sandoval, J., Monk, D., Hata, K., Marques-Bonet, T. and Wang, L. (2013) DNA methylation contributes to natural human variation. *Genome Res.*, **23**, 1363–1372.
  90. Barfield, R.T., Almli, L.M., Kilaru, V., Smith, A.K., Mercer, K.B., Duncan, R., Klengel, T., Mehta, D., Binder, E.B. and Epstein, M.P. (2014) Accounting for population stratification in DNA methylation studies. *Genet. Epidemiol.*, **38**, 231–241.
  91. Fraser, H.B., Lam, L.L., Neumann, S.M. and Kobor, M.S. (2012) Population-specificity of human DNA methylation. *Genome Biol.*, **13**, R8.
  92. Gopalan, S., Carja, O., Fagny, M., Patin, E., Myrick, J.W., McEwen, L.M., Mah, S.M., Kobor, M.S., Froment, A. and Feldman, M.W. (2017) Trends in DNA methylation with age replicate across diverse human populations. *Genetics*, **206**, 1659–1674.
  93. Kooijman, M.N., Kruithof, C.J., van Duijn, C.M., Duijts, L., Franco, O.H., van Ijzendoorn, M.H., de Jongste, J.C., Klaver, C.C.W., van der Lugt, A. and Mackenbach, J.P. (2016) The generation R study: design and cohort update 2017. *Eur. J. Epidemiol.*, **31**, 1243–1264.
  94. Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. and Davey Smith, G. (2013) Cohort profile: the ‘children of the 90s’—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.*, **42**, 111–127.
  95. Medina-Gomez, C., Felix, J.F., Estrada, K., Peters, M.J., Herrera, L., Kruithof, C.J., Duijts, L., Hofman, A., van Duijn, C.M. and Uitterlinden, A.G. (2015) Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the generation R study. *Eur. J. Epidemiol.*, **30**, 317–330.
  96. Relton, C.L., Gaunt, T., McArdle, W., Ho, K., Duggirala, A., Shihab, H., Woodward, G., Lyttleton, O., Evans, D.M. and Reik, W. (2015) Data resource profile: accessible resource for integrated epigenomic studies (ARIES). *Int. J. Epidemiol.*, **44**, 1181–1190.
  97. Lehne, B., Drong, A.W., Loh, M., Zhang, W., Scott, W.R., Tan, S.-T., Afzal, U., Scott, J., Jarvelin, M.-R. and Elliott, P. (2015) A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.*, **16**, 37.
  98. Min, J.L., Hemani, G., Davey Smith, G., Relton, C., Suderman, M. and Hancock, J. (2018) Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics*, **34**, 3983–3989.
  99. R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.



100. Bakulski, K.M., Feinberg, J.I., Andrews, S.V., Yang, J., Brown, S., L. McKenney, S., Witter, F., Walston, J., Feinberg, A.P. and Fallin, M.D. (2016) DNA methylation of cord blood cell types: applications for mixed cell birth studies. *Epigenetics*, **11**, 354–362.
101. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K. and Kelsey, K.T. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, **13**, 1.
102. Hothorn, T., Bretz, F., Westfall, P., Heiberger, R.M., Schuetzenmeister, A., Scheibe, S. and Hothorn, M.T. (2016) Package 'multcomp'. *Simultaneous inference in general parametric models*. Project for Statistical Computing, Vienna, Austria.
103. Bates, D., Mächler, M., Bolker, B. and Walker, S. (2018) *Package lme4. Version 1*, 17.
104. Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S. and Gonzalez, J.N. (2019) The UCSC genome browser database: 2019 update. *Nucleic Acids Res.*, **47**, D853–D858.
105. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C. and Suzuki, T. (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455–461.
106. Cecil, C.A.M., Smith, R.G., Walton, E., Mill, J., McCrory, E.J. and Viding, E. (2016) Epigenetic signatures of childhood abuse and neglect: implications for psychiatric vulnerability. *J. Psychiatr. Res.*, **83**, 184–194.