

# Childhood environment influences epigenetic age and methylation concordance of a CpG clock locus in British-Bangladeshi migrants

Reinhard Stöger, Minseung Choi, Khurshida Begum, Gregory Leeman, Richard D. Emes, Philippa Melamed & Gillian R. Bentley

To cite this article: Reinhard Stöger, Minseung Choi, Khurshida Begum, Gregory Leeman, Richard D. Emes, Philippa Melamed & Gillian R. Bentley (2022): Childhood environment influences epigenetic age and methylation concordance of a CpG clock locus in British-Bangladeshi migrants, Epigenetics, DOI: [10.1080/15592294.2022.2153511](https://doi.org/10.1080/15592294.2022.2153511)

To link to this article: <https://doi.org/10.1080/15592294.2022.2153511>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 10 Dec 2022.



[Submit your article to this journal](#)





[View related articles](#)









[View Crossmark data](#)

RESEARCH PAPER

 OPEN ACCESS 

## Childhood environment influences epigenetic age and methylation concordance of a CpG clock locus in British-Bangladeshi migrants

Reinhard Stöger <sup>a</sup>, Minseung Choi <sup>b</sup>, Khurshida Begum <sup>c</sup>, Gregory Leeman<sup>a</sup>, Richard D. Emes <sup>d,e</sup>, Philippa Melamed <sup>f</sup>, and Gillian R. Bentley <sup>c,g</sup>

<sup>a</sup>School of Biosciences, University of Nottingham, Nottingham, UK; <sup>b</sup>School of Medicine, Stanford University, Stanford, CA, USA; <sup>c</sup>Department of Anthropology, Durham University, Durham, UK; <sup>d</sup>School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK; <sup>e</sup>Advanced Data Analysis Centre, University of Nottingham, Nottingham, UK; <sup>f</sup>Faculty of Biology, Technion-Israel Institute of Technology, Haifa, Israel; <sup>g</sup>Wolfson Research Institute for Health and Wellbeing, Durham University, Durham, UK

### ABSTRACT

Migration from one location to another often comes with a change in environmental conditions. Here, we analysed features of DNA methylation in young, adult British-Bangladeshi women who experienced different environments during their childhoods: *a*) migrants, who grew up in Bangladesh with exposure to comparatively higher pathogen loads and poorer health care, and *b*) second-generation British-Bangladeshis, born to Bangladeshi parents, who grew up in the UK. We used buccal DNA to estimate DNA methylation-based age (DNAm age) from 14 migrants and 11 second-generation migrants, aged 18–35 years. ‘AgeAccel,’ a measure of DNAm age, independent of chronological age, showed that the group of women who spent their childhood in Bangladesh had higher AgeAccel ( $P = 0.028$ ), compared to their UK peers. Since epigenetic clocks have been proposed to be associated with maintenance processes of epigenetic systems, we evaluated the preference for concordant DNA methylation at the luteinizing hormone/choriogonadotropin receptor (*LHCGR/LHR*) locus, which harbours one of the CpGs contributing to Horvath’s epigenetic clock. Measurements on both strands of individual, double-stranded DNA molecules indicate higher stability of DNA methylation states at this *LHCGR/LHR* locus in samples of women who grew up in Bangladesh. Together, our two independent analytical approaches imply that childhood environments may induce subtle changes that are detectable long after exposure occurred, which might reflect altered activity of the epigenetic maintenance system or a difference in the proportion of cell types in buccal tissue. This exploratory work supports our earlier findings that adverse childhood environments lead to phenotypic life history trade-offs.

### ARTICLE HISTORY

Received 18 October 2021  
Revised 16 Nov 2022  
Accepted 23 November 2022

### KEYWORDS

Childhood environment; migrants; AgeAccel; methylation concordance; epigenetic clock; Bangladesh; UK


## Background

Biological ageing, the waning of physiological functions over time is inherent to eukaryotes, including humans. Individuals age at different rates and this affects their health and life expectancy [1]. Twin studies suggest a moderate genetic involvement of about 25% in human lifespan variation [2–4]; an influence that becomes apparent mainly in later life past the age of 60 [2]. Stochastic processes may also cause heterogeneity in the rate at which people age [5], but environmental factors appear to have the strongest impact on human life expectancy, as shown by census data and population studies [6]. During the last seven decades, the global average life expectancy increased by about 22 years for both men (70.5 years) and women (75.6 years) [7]. The remarkable increase in

lifespan is attributed to advances in hygiene, nutrition, and health care which, in concert, have significantly reduced child mortality [8]. Yet, substantial health disparities among populations persist, many of which are linked to specific geographic influences [9] that impact on life history traits, disability adjusted life years, and lifespan [10,11]. Investigators are now using molecular biomarkers, such as telomere length and epigenetic age, to gain a better understanding of the impact of environmental factors on morbidity and mortality [1,12,13].

Environmental factors influence cell-type composition of tissues [14] and chromatin states within cells [15]. Using epigenetic information, cells have the capacity to retain some memory of past developmental and environmental conditions [16]. Methylation

**CONTACT** Reinhard Stöger  reinhard.stoger@nottingham.ac.uk  School of Biosciences, University of Nottingham, Nottingham, UK

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15592294.2022.2153511>

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

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 use, distribution, and reproduction in any medium, provided the original work is properly cited.

levels of discrete CpG sites have been used to develop remarkably accurate estimators of chronological age. Such ‘epigenetic clocks’ have been proposed to link developmental and maintenance processes with biological ageing [17–19]. Biological ageing varies across individuals, at times resulting in a mismatch between chronological and biological age [20].

Environmental conditions experienced during early life influence individuals in ways, which may affect their health during adulthood [21]. The developing body may invest limited resources towards physiological demands and biological fitness traits that enable survival during early life at the expense of other characteristics in later life [22]. Such life history ‘trade-offs’ and interactions of fitness traits include growth patterns, timing of puberty, immune function and, in adulthood, characteristics such as age at menopause and lifespan [23–26].

Our earlier work has already identified consistent associations between childhood environmental conditions and the adult phenotype [27–29]. Bangladeshi women who migrated as adults (defined as >16 years old) to London, had lower levels of salivary progesterone, a later age at puberty and earlier ages at menopause when compared to British-Bangladeshi women who moved as children to the UK and second-generation women born in London to first-generation Bangladeshi immigrants [28–31]. Age at adrenarche was also earlier among recent child migrants to the UK [31]. Geographically and culturally, the British-Bangladeshis in these studies have a comparable background and ancestry. They are all ethnic Bengalis, are mostly Muslim, and originate from a relatively affluent middle-class population living in Sylhet District in the northeast of Bangladesh [32]. Once in the UK, they traditionally settled in ethnic enclaves in east London and other neighbourhoods [33,34] but this pattern is now changing as people move out to further suburbs [35]. A childhood spent in Bangladesh is generally associated with a shortened reproductive lifespan [36] relative to women with Bangladeshi ancestry whose childhood was spent in London [28]. A possible environmental factor that distinguishes between the two childhood locations is exposure to higher and recurrent infectious disease loads in Bangladesh [36–38]. Indeed, by mimicking early-life immune

challenges in a mouse model, we have replicated some of the distinct reproductive phenotypes characteristic of women with a childhood in Bangladesh, including a delayed onset of puberty and lower ovarian reserve [28,36,39].

In the current study, we explored *i*) the possible association between chronological age, biological ageing and childhood location, as well as *ii*) the stability of the epigenetic maintenance system at one location within the genome in Bangladeshi women, aged between 18 and 35 years. The women in this study lived within the same ethnic community in London but were divided into two groups: *a*) those with a childhood spent in Sylhet, who immigrated to the UK when they were >16 years of age, and *b*) those with a childhood spent in the UK (that is, born in the UK or immigrated < 2 years of age). By profiling buccal cell DNA on the Illumina 850 MethylationEpic array platform, we previously identified genome-wide, altered DNA methylation levels between the two groups of London-based Bangladeshi women [39]. Here, we used these methylation data to establish epigenetic age predictions, with three different epigenetic clocks: ‘Horvath’s multi-tissue’ clock, a ‘skin & blood’ clock, and a ‘3-CpG-swab’ model [40–42]. The tick rate of Horvath’s epigenetic clock is thought to reflect the rate at which work is undertaken to maintain epigenetic stability within a cell [17,41]. It is possible to infer epigenetic stability with a new metric, the Ratio of Concordance Preference (RCP), by analysing methylation data derived from individual, double-stranded DNA molecules [43]. We generated such data from the *LHCGR/LHR* locus to probe the efficiency of maintaining the DNA methylation state, perhaps capturing a minuscule aspect of what epigenetic clocks may register [17,18].

## Results and discussion

### *Accelerated DNAm age measured with Horvath’s epigenetic clock*

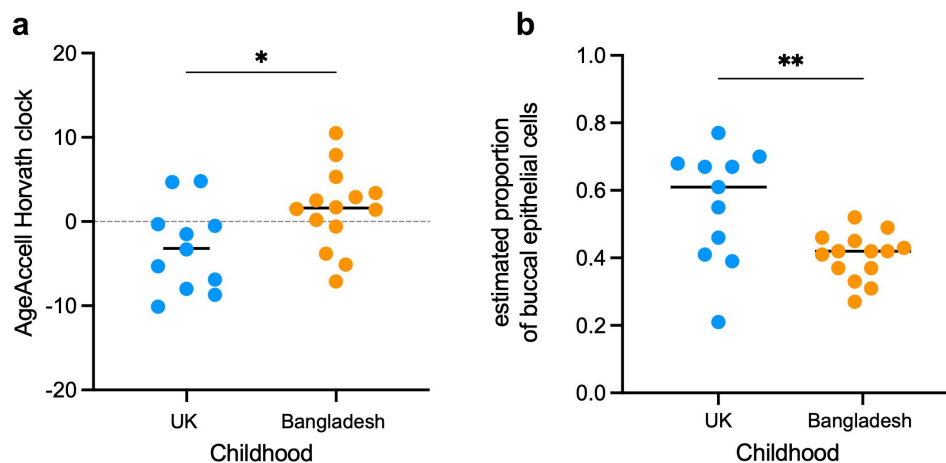
We found that the correlation between chronological age and DNAm age did not differ significantly between women who grew up in the UK (‘UK’ group;  $n = 11$ ) and women who grew up in Bangladesh (‘Bangladesh’ group;  $n = 14$ ). That is,

chronological age affected DNAm age in similar ways in both groups (Additional File: Figure S1 & Table 1). However, regression analysis indicated that the y-intercepts of the UK and Bangladeshi groups did differ significantly ( $P = 0.009$ ) which suggested divergent DNAm age predictions between the two cohorts (Additional File: Figure S1 & Table 1). Potential differences in epigenetic age estimates and chronological age, including accelerated epigenetic ageing, are best measured by comparing the distribution of residuals obtained from a linear model, where DNAm age is regressed on chronological age (AgeAccel). Such DNAm age AgeAccel biomarkers provide epigenetic age estimates that are independent of chronological age: a positive AgeAccel value indicates that the epigenetic age is higher than expected, whereas a negative value indicates a lower than expected epigenetic age relative to chronological age [41]. With Horvath's multi-tissue clock [41] this approach revealed significant AgeAccel ( $P = 0.028$ ) for women with a childhood in Bangladesh, suggesting an accelerated rate of epigenetic ageing [AgeAccel mean: +1.5 (Figure 1a)]. In contrast, a childhood in the UK was associated with decreased AgeAccel, implying that these women experience slower rates of epigenetic ageing [AgeAccel mean: -3.2 (Figure 1a)]. No

significant differences were observed between the 'UK' and the 'Bangladeshi' groups with AgeAccel biomarkers derived from the 'skin & blood' clock and the '3-CpG-swab' model (Additional File: Table 2 & Figure S2).

Our finding of epigenetic age differences based on Horvath's multi-tissue clock between women associated with two different geographic childhood environments rests on a small number of sampled individuals. We interpret this result with caution and our finding will be strengthened by independent validation and larger datasets.

The composite nature of buccal tissue could, in part, explain some of the variation in our AgeAccel measurements. In fact, single-cell RNA sequencing revealed the presence of at least eight different subtypes of cells in the buccal mucosal basal layer; turnover of these cells follows a pattern of neutral drift dynamics and is influenced by age and environmental factors [44]. The range of individual DNAm age estimates from buccal epithelial cells is indeed broader, when compared with intra-individual DNAm age estimates from blood-cells, as Hannon and colleagues found, when they assessed Horvath's epigenetic clock across tissues and cells [45]. Nevertheless, the same study showed that the mean of these epigenetic age



**Figure 1.** Differences in pace of epigenetic ageing and buccal cell composition.

**a)** Plot of DNAm AgeAccel residuals derived from Horvath's multi tissue clock, where each data point represents a value from an individual. Colours identify the corresponding dataset: blue = childhood in UK ( $n = 11$ ), orange = childhood in Bangladesh ( $n = 14$ ). The median is indicated by a black horizontal line. A positive or negative AgeAccel value indicates that the estimated epigenetic/biological age of the sample is higher or lower, respectively, than expected. **b)** The estimated fraction of epithelial cells in buccal samples based on a 2-CpG buccal cell signature model [40]. Asterisks denote two-tailed significance ( $* = P \leq 0.05$ ;  $** P \leq 0.01$ ) between groups, based on unpaired t-tests.

predictions was more accurate for buccal samples than for blood samples [45]. We therefore considered the composition and proportion of cells in our buccal samples. Deconvolution methods for DNA methylation data have been developed that provide estimates of the abundance of certain cell subtypes [46,47]. The tissue predicting tool, integrated with the online methylation age calculator [48], determined ‘buccal/saliva’ as the most likely tissue source for all the DNA samples described here (Additional File: Table 2). Wagner and colleagues developed a model based on methylation levels of two CpG sites, which estimates the proportions of buccal epithelial cells and leukocytes in buccal swab samples [40]. With its ‘buccal-cell-signature’ model, we observed a significant difference ( $P = 0.006$ ) in epithelial cell composition between the UK and Bangladeshi samples (Figure 1b). The overall lower proportions of epithelial cells estimated to be present in buccal swabs from women with a Bangladeshi childhood cannot be easily attributed to differences in collection and processing. Buccal swabs were taken randomly from individuals with UK and Bangladeshi childhoods, but always under supervision of the same person, ensuring consistency of the collection procedure (see also Additional File: Figure 3).

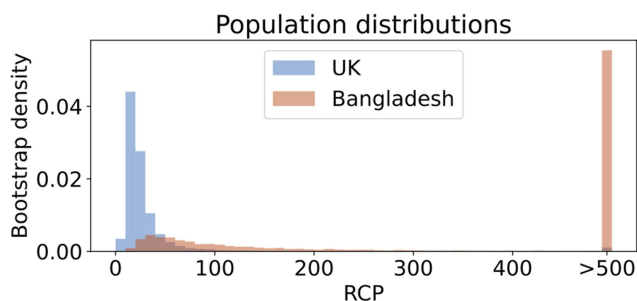
Altered hormone levels could also have contributed to the variation and differences in AgeAccel between the two cohorts. Reproductive hormone levels such as oestrogen and progesterone appear to influence measures of AgeAccel obtained from buccal epithelium, in contrast to blood-derived AgeAccel estimates [49]. That is, menopausal hormone therapies (MHT) are associated with significantly lower epigenetic AgeAccel in buccal tissue [49]. While it is unlikely that the women in our study, aged between 18 and 35 years, were undergoing MHT, we cannot rule out the possibility that some might have used oral contraceptives at the time of buccal sample collection. Alternatively, we could argue that the childhood environment has led to comparatively higher levels in reproductive hormones in the UK group of women, as we previously reported [36], resulting in decreased epigenetic AgeAccel. Future epigenetic age studies

using buccal DNA samples will benefit from integrating additional information about the participants.

Duration, intensity and the type of exposure experienced during early life may influence methylation of some clock-associated CpG sites [50]. At this stage, we know that immunological challenges play a role [30,39], but defining the timeframe of exposures during early and middle childhood will be critical to understand complex phenotypic outcomes that appear to be associated with the tick rate of the epigenetic clock. Accelerated epigenetic ageing is consistent with our previous observations of shorter reproductive lifespans and chronically lower levels of reproductive hormones in women who grew up in Bangladesh [28,36,38] (reviewed in [30]), as well as with reduced oocyte yields at the time of retrieval in women undergoing fertility evaluation [51].

### Epigenetic stability of a clock locus

We used the RCP metric [43] to estimate epigenetic stability in samples, where sufficient DNA was available (Bangladeshi-born,  $n = 6$  / UK-born,  $n = 6$ ). Specifically, a  $\sim 100$  bp sequence of the *Luteinizing Hormone/Choriogonadotropin Receptor (LHCGR/LHR)* locus was analysed, which contains one of the CpG sites comprising Horvath’s multi-tissue epigenetic clock, along with flanking CpG sites (Additional File: Figure 6 and Table 3). This clock-associated CpG site was chosen because *LHCGR/LHR* is necessary for ovulation and luteal formation in females [52,53], and its epigenetic state could be associated with differences in reproductive function, a phenotype we previously linked with altered environmental conditions [28,30,36,54]. We found that RCP estimates were generally higher for the adult migrants who spent their childhoods in Bangladesh (Figure 2). Higher RCP estimates indicate higher levels of epigenetic stability [43]. That is, the methylation states of five or eight CpG dyads (see Additional File: Figure 6), at this *LHCGR*-associated locus are more often identical on the two strands of individual DNA molecules of adult migrants compared with those inferred from DNA samples of ‘second generation’ Bangladeshi women, who grew up in the UK. We note that the RCP estimates are based on a relatively small number of data points (Additional File:



**Figure 2.** Inferences of DNA methylation stability at the epigenetic-clock associated *LHCGR/LHR* locus.

The analysis is based on a sampled population of double-stranded DNA molecules and the corresponding methylation states at matching CpG sites (CpG dyads) on the parent and daughter strand of individual molecules. Ratio of Concordance Preference (RCP) is a metric that infers stability/flexibility of methylation states at CpG dyads, without assuming any specific enzymatic mechanisms of DNA methylation. Flexibility, indicated by RCP values near 1, imply that the methylation system has no preference for either concordance or discordance of the methylation state at a CpG dyad and is distributed as expected under a null model, in which methyl groups are placed at random on both strands of a DNA molecule. High RCP values – with the extreme approaching infinity – indicate high stability, where epigenetic maintenance systems have complete preference for concordant methylation states of CpG dyads (they are either methylated or unmethylated); none, or very few CpG dyads are hemi-methylated. Shown are the RCP distributions taken from bootstrap samples, weighing each individual evenly within each group. The sampled population of double-stranded DNA molecules – and the corresponding methylation states of CpG dyads – revealed that the epigenetic maintenance system operates with a tendency of increased stability at the *LHCGR/LHR* locus in women with a childhood in Bangladesh, when compared to the group of women with a childhood in London/UK.

Figure 4 and Figure 5), derived from only a single genomic locus. The results are consistent with the idea of what clock-CpGs might proxy: the workings of the epigenetic maintenance system [17]. In our case, subtle differences between the two groups of women is inferred by RCP analysis. Whether variation in the efficiency of the methylation maintenance system is locus-specific, or a genome-wide phenomenon remains to be demonstrated by analysing additional genomic loci or using data obtained from high-resolution deep hairpin bisulphite sequencing (DHBS) [55,56].

## Conclusions

The results of our study support a large body of work demonstrating phenotypic plasticity in response to

environments encountered during early life. A childhood in Bangladesh appears to be associated with a faster rate of epigenetic/biological ageing in young adult females, when compared to women of a similar chronological age (18–35 years) and cultural background, who were born and brought up in London, UK. The multi-tissue epigenetic clock is thought to register the workings of developmental and epigenetic maintenance systems linking these processes across the life course [17,41]. Along with work by Martin-Herranz and colleagues [57], this study is one of the first to explore if differences in function of the epigenetic maintenance system can be linked with epigenetic age estimators. Our RCP measurements detected subtle differences in the stability of epigenetic states, which could be a result of minute changes in the activity of the DNA methylation maintenance system. Alternatively, the RCP measurements may have detected altered DNA methylation stabilities that resulted from differences in the proportions of dividing buccal cell subtypes, leading in sum to dissimilar epigenomes in our two groups of DNA samples.

## Methods

### Collection of buccal cell samples

Buccal cell samples were collected from donors of both, the ‘UK group’ and the ‘Bangladeshi group’ within a ~ two-month period (May–June 2016), all under supervision by KB, thereby ensuring consistency in the process. Swabs were placed in an iSWAB™ collection device containing a proprietary lysis and stabilizing buffer following the instructions of the iSWAB DNA Collection Kit (Mawi DNA Technologies). The lysed buccal cell samples from all individuals were stored under same conditions until further processing.

### DNA methylation data, establishment of DNAm Age and AgeAccel

Genome-wide cytosine methylation levels were established using the Illumina HumanMethylationEPIC BeadChip Array following isolation of genomic DNA from the collected and lysed buccal cell samples (DNeasy Blood & Tissue Kit (Qiagen). Multidimensional scaling (MDS) plots indicated that no

significant batch effects were skewing the MethylationEPIC BeadChip data sets. The data were processed with the Bioconductor/minfi package. CpG probes associated with known SNPs were removed, as were those with a detection probability of  $<0.01$ . Probes on both X and Y chromosomes were retained. Methylation beta values (0–1) were analysed using the Horvath online multi-tissue epigenetic clock [48] and the skin-blood clock [42] following two normalization procedures (SWAN, Quantile) or no-normalization (raw data); for the ‘3-CpG model’ [40] only raw (no normalization) data were used. The methylation data set (GSE133355 study) is accessible on the Gene Expression Omnibus (GEO) data platform at: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE133355>

Files with the beta values obtained from the Illumina HumanMethylationEPIC BeadChip Array work (see above) were used to establish the epigenetic age (DNAm age and AgeAccelerationResiduals – AgeAccel) with Horvath’s method [41]. The underlying algorithms are available through the online DNAm age calculator (<https://dnamage.genetics.ucla.edu/new>). We used the default setting ‘Normalize Data’ during the online-submission process to this DNAm Age calculator as it often improves the predictive accuracy, since this normalization method renders the input data comparable to the training data of the ‘Horvath clock.’ Output from this online calculator also gives predictions of sex and tissue/cell source of the samples, and this confirmed that all our samples described here were indeed female and the source of tissue was buccal/saliva.

### Generation of bisulphite hairpin data/ methylation states of CpG dyads at the LHCGR/LHR locus

We have previously described in detail the concept and procedure of generating authenticated, non-redundant double-stranded DNA methylation data [58–60]. In brief, genomic sequence information surrounding the LHR clock-CpG site [one of the 353 CpG sites contributing to Horvath’s clock [41]. Illumina cluster ID cg12351433/chr2:48,982,957–48,982,957/UCSC Genome Browser (GRCh37/hg19)] was used to identify suitable restriction recognition sites to generate 3’, or 5’-overhangs, respectively, for the ligation

of UMI-barcoded hairpin linkers. Specifically, restriction enzymes StyI or BstXI (New England Biolabs) were used. Combinations of the following primers were used to amplify hairpin-linked, bisulphite converted DNA:

bsLHR-R1 5’-RCAAATCAAAACAAAACAAAC TC-3’;

bsLHR-R2 5’-CACTAAACACTATCRCAAATCA AAAC-3’;

bsLHR-F1 5’-TAGTAGGAAGGAGGTTATTGG-3’;

bsLHR-F2 5’-GTAGGTTAAGGTAGAGTAGATTT AG-3’;

bsLHR-F3 5’-GAATTGGGTTTTTGC GGTTTGTT AG-3’.

Further information of the hairpin-concept and of the barcoded and batch-stamped hairpin linkers (Eurofins Genomics) are provided in Additional File 1.

### Processing of the sequencing data:

Fold is a web application for the analysis of the output of hairpin-bisulphite sequencing data. Specifically, the programme reconstructs, visualizes, and generates statistics on the double-stranded CpG methylation patterns of the original cohort of DNA molecules. This is achieved by first ‘realigning’ the top and bottom strand of the molecule about the hairpin, in which the programme attempts to manage ‘PCR slippage,’ and other sequencing errors. The algorithm then identifies and categorizes CpG dyads, which is possible due to the previous bisulphite conversion of unmethylated cytosine to uracil (and so recognized as tyrosine when sequenced). For example, fully methylated dyads are those regions where the reconstructed top strand is C-G and the bottom is G-C. Similarly, fully unmethylated dyads are those where the top is T-G and the bottom is G-T. In addition, the programme calculates a metric: ‘Ratio of Concordance Preference’ which quantifies the concordance of methylation between the top and bottom strands of the DNA molecule (0 = complete discordance, 1 = random concordance,  $\infty$  = complete concordance). This metric represents the preference of the summation of epigenetic mechanisms of the cell to either maintain or obscure methylation patterns of the DNA in the cells at the time the sample was taken.

The functions of Fold were written in R and the web application is written in PHP. The live web application can be found at <http://www.gregoryleeman.com/fold>, and the repository can be found at <https://github.com/gregoryleeman/fold>.

### Analysis and comparison of RCPs at the LHCGR/LHR locus

RCP values are based on double-stranded DNA methylation data derived from sequences of individual hairpin bisulphite PCR products. RCP is defined as:

$$RCP = \frac{\sqrt{U(U + 2m - 1)}}{1 - U - m}$$

RCP analyses were undertaken following the procedures described in [43], with the small addition of bootstrapping individuals within each population. The additional step in the procedure helps to address the possibility of uneven sampling from a larger population of individuals. The analysis procedures in brief are described below.

Each population RCP distribution was drawn through hierarchical bootstrap sampling. For each of 20,000 bootstrap samples, individuals in each population were sampled with replacement, and double stranded DNA sequences of each of the sampled individuals were in turn sampled with replacement. Dyad counts were then normalized, such that each individual had the same number of dyads. The normalized dyad counts were then summed, corrected for failed bisulphite conversions (rate of 0.0039, measured empirically) and inappropriate conversions (rate of 0.017, estimated as described in [43]) and used to compute the RCP value. A bootstrap sample of the RCP difference was computed by taking the difference of the RCP values sampled for the two populations.

Assuming that there are no locus-specific differences between the two groups, differences in numbers of CpG Dyads should not affect the results of RCP analysis. RCP treats each dyad independently and does not rely on sequential aspects of the dyads in the data. The number of CpG sites per hairpin, however, does influence the variance of the data, as sites on a single strand are correlated. We take this into account in our hierarchical bootstrapping approach.

For one-tailed comparison tests, with which we examine directional differences, we determined the p-value as the proportion of bootstrap-difference samples to the left of 0. For two-tailed tests, with which we can detect differences in any direction, we determined the p-value as twice the smaller proportion of the bootstrap difference samples on either side of 0.

### Acknowledgments

We thank Kamila Derecka for technical support, Ian C. W. Hardy for statistical advice and Steve Horvath for information on DNAm age analysis, Wolfgang Wagner for help in analysing the data with the 3-CpG model and the ‘Buccal-Cell-Signature’ estimates, Diane Genereux and Charles Laird for comments on the manuscript.

### Authors’ contributions

RS: Conceptualisation, Funding acquisition, Experimental work, Analysis, Resources, Supervision, Data curation, Project administration, Writing – original draft, Writing – review & editing.

MC: Analysis, review & editing

GL: Analysis, Coding

RDE: Analysis, Data curation

KB: Experimental work, Resources

PM: Funding acquisition, Project administration, Writing – review & editing.

GRB: Conceptualisation, Funding acquisition, Project administration, Writing – review & editing.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

This research was supported by the Economic and Social Research Council (ESRC) and the Biotechnology and Biological Science Research Council (BBSRC) grant ES/N000471/1(to GB, PM and RS).

### Ethics

Protocols for human data collection were approved by the Ethical Committee of the Department of Anthropology, Durham University. Data were anonymized at source.



## Data availability statement

The datasets generated and/or analysed during the current study are available in the Gene Expression Omnibus (GEO) data platform <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE133355>

All data generated or analysed during this study are included in this published article and its supplementary information file.

## ORCID

Reinhard Stöger  <http://orcid.org/0000-0002-2301-0805>  
 Minseung Choi  <http://orcid.org/0000-0001-5398-219X>  
 Khurshida Begum  <http://orcid.org/0000-0001-9465-1100>  
 Richard D. Emes  <http://orcid.org/0000-0001-6855-5481>  
 Philippa Melamed  <http://orcid.org/0000-0001-7814-6542>  
 Gillian R. Bentley  <http://orcid.org/0000-0002-2009-5540>

## References

- [1] Jylhävä J, Pedersen NL, Hägg S. Biological age predictors. *EBioMedicine* [Internet]. 2017. [cited 2018 Jul 4];21:29–36.
- [2] Hjelmborg JB, Iachine I, Skytthe A, et al. Genetic influence on human lifespan and longevity. *Hum Genet* [Internet]. 2006[cited 2020 Nov 16]; 119:312–321. <https://link.springer.com/article/10.1007/s00439-006-0144-y>
- [3] Martin GM, Bergman A, Barzilai N. Genetic determinants of human health span and life span: progress and new opportunities. *PLoS Genet* [Internet]. 2007[cited 2020 Nov 16]; 3:e125. <https://dx.plos.org/10.1371/journal.pgen.0030125>
- [4] Slagboom PE, Beekman M, Passtoors WM, et al. Genomics of human longevity. *Philos Trans R Soc B* [Internet]. 2011. [cited 2020 Nov 16];366:35–42. <https://royalsocietypublishing.org/doi/10.1098/rstb.2010.0284>
- [5] Martin GM. Stochastic modulations of the pace and patterns of ageing: impacts on quasi-stochastic distributions of multiple geriatric pathologies. *Mech Ageing Dev* [Internet]. 2012 [cited 2020 Nov 16];133:107–111.
- [6] Eline Slagboom P, van den Berg N, Deelen J. Phenome and genome based studies into human ageing and longevity: an overview. *Biochim Biophys Acta Mol Basis Dis*. 2018;1864:2742–2751.
- [7] Dicker D, Nguyen G, Abate D, et al. Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the global burden of disease study 2017. *Lancet*. 2018;392:1684–1735.
- [8] Moser K, Shkolnikov V, Leon DA. World mortality 1950–2000: divergence replaces convergence from the late 1980s. *Bull World Health Organ*. 2005;83:202–209.
- [9] Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet* [Internet]. 2020. [cited 2021 Aug 31];396:1204–1222. <http://www.thelancet.com/article/S0140673620309259/fulltext>
- [10] Nettle D. Dying young and living fast: variation in life history across English neighborhoods. *Behav Ecol* [Internet]. 2010 [cited 2021 Aug 31];21:387–395. <https://academic.oup.com/beheco/article/21/2/387/321953>
- [11] Marmot M, Wilkinson RG. *Social Determinants of Health*. Oxford United Kingdom: Oxford University Press; 2009.
- [12] Tennyson RL, Gettler LT, Kuzawa CW, et al. Lifetime socioeconomic status and early life microbial environments predict adult blood telomere length in the Philippines. *American J Human Bio* [Internet] 2018 cited 2021 Sep 3; 30:e23145. <https://onlinelibrary.wiley.com/doi/full/10.1002/ajhb.23145>.
- [13] Zalli A, Carvalho LA, Lin J, et al. Shorter telomeres with high telomerase activity are associated with raised allostatic load and impoverished psychosocial resources. *Proceedings of the National Academy of Sciences* [Internet] 2014 cited 2021 Sep 3; 111:4519–4524. <https://www.pnas.org/content/111/12/4519>
- [14] Lappalainen T, Grealley JM. Associating cellular epigenetic models with human phenotypes. *Nat Rev Genet* [Internet]. 2017 [cited 2019 Jul 18];18:441–451. <http://www.nature.com/articles/nrg.2017.32>
- [15] Zhu J, Adli M, Zou JY, et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*. 2013 [cited 2019 Sep 4]; 152:642–654. <http://www.ncbi.nlm.nih.gov/pubmed/23333102>
- [16] Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* [Internet]. 2003;33 (Suppl):245–254. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=12610534](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12610534)
- [17] Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* [Internet]. 2018 [cited 2018 Nov 15];19:371–384. <http://www.nature.com/articles/s41576-018-0004-3>
- [18] Field AE, Robertson NA, Wang T, et al. DNA methylation clocks in aging: categories, causes, and consequences. *Mol Cell* [Internet]. 2018 [cited 2019 Sep 10];71:882–895. <http://www.ncbi.nlm.nih.gov/pubmed/30241605>
- [19] Simpson DJ, Chandra T. Epigenetic age prediction. *Aging Cell*. 2021 [cited 2021 Sep 17];20:e13452. <https://onlinelibrary.wiley.com/doi/full/10.1111/accel.13452>
- [20] Khan SS, Singer BD, Vaughan DE. Molecular and physiological manifestations and measurement of

- aging in humans. *Aging Cell*. 2017 [cited 2020 Jun 8];16:624–633. <http://doi.org/10.1111/acel.12601>
- [21] Gluckman PD, Hanson MA, Beedle AS. Early life events and their consequences for later disease: a life history and evolutionary perspective. *American J Human Bio* [Internet] 2007 cited 2021 Sep 3; 19:1–19. <https://onlinelibrary.wiley.com/doi/full/10.1002/ajhb.20590>.
- [22] Worthman CM, Kuzara J. Life history and the early origins of health differentials. *Am J Hum Biol* [Internet]. 2005;17:95–112. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=15611966](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15611966)
- [23] Hill K. Life history theory and evolutionary anthropology. *Evolutionary Anthropology: issues, News, and Reviews* [Internet] 1993 cited 2021 Sep 3; 2:78–88. <https://onlinelibrary.wiley.com/doi/full/10.1002/evan.1360020303>.
- [24] Roff DA. *The evolution of life histories: theory and analysis*. 1st ed. New York:Springer;1993.
- [25] Stearns SC. *The evolution of life histories*. Oxford United Kingdom: Oxford University Press; 1992.
- [26] Lochmiller RL, Deerenberg C. Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos* [Internet]. 2000[cited 2020 Nov 22]; 88:87–98. <https://onlinelibrary.wiley.com/doi/full/10.1034/j.1600-0706.2000.880110.x>
- [27] Magid K, Chatterton RT, Ahamed FU, et al. Childhood ecology influences salivary testosterone, pubertal age and stature of Bangladeshi UK migrant men. *Nat Ecol Evol* [Internet]. 2018 [cited 2019 Sep 6];2:1146–1154. <http://www.nature.com/articles/s41559-018-0567-6>
- [28] Begum K, Muttukrishna S, Sievert L, et al. Ethnicity or environment: effects of migration on ovarian reserve among Bangladeshi women in the UK. *Fertil Steril*. 2016;105:744–754.
- [29] Nunez-De La Mora A, Bentley GR, Choudhury OA, et al. The impact of developmental conditions on adult salivary estradiol levels: why this differs from progesterone? *Am J Hum Biol* [Internet]. 2008;20:2–14. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=17957763](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17957763)
- [30] Bar-Sadeh B, Rudnizky S, Pnueli L, et al. Unravelling the role of epigenetics in reproductive adaptations to early-life environment. *Nat Rev Endocrinol* [Internet]. 2020 [cited 2020 Aug 18];16:519–533. <https://www.nature.com/articles/s41574-020-0370-8>
- [31] Houghton LC, Cooper GD, Booth M, et al. Childhood environment influences adrenarcheal timing among first-generation Bangladeshi migrant girls to the UK. *PLoS One*. 2014 [cited 2019 Sep 6];9:e109200. <https://dx.plos.org/10.1371/journal.pone.0109200>
- [32] Gardner K. International migration and the rural context in Sylhet. *New community* [Internet] 2010 cited 2021 Sep 3; 18:579–590. <https://www.tandfonline.com/doi/abs/10.1080/1369183X.1992.9976331>.
- [33] Carey S, Shukur A. A profile of the Bangladeshi community in East London. 2010;New community. 12:405–417.
- [34] Eade J, Vamplew T, Peach C. The Bangladeshis: the encapsulated community. In: Peach C, editor. *The ethnic minority populations of Great Britain: volume 2 of ethnicity in the 1991 census*. London: Office for National Statistics | HMSO; 1996. p. 150–160.
- [35] Stidle L, Corporate\_Research\_Unit. *A profile of the migrant population in tower hamlets*. London; 2017.
- [36] Nunez-de la Mora A, Chatterton RT, Choudhury OA, et al. Childhood conditions influence adult progesterone levels. *PLoS Med* [Internet]. 2007 [cited 2019 Sep 6];4:e167. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=17503960](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17503960)
- [37] Siddique AK, Baqui AH, Eusof A, et al. 1988 floods in Bangladesh: pattern of illness and causes of death. *J Diarrhoeal Dis Res* [Internet]. 1991[cited 2020 Sep 17]; 9:310–314. <https://pubmed.ncbi.nlm.nih.gov/1800561/>
- [38] Murphy L, Sievert L, Begum K, et al. Life course effects on age at menopause among Bangladeshi sedentees and migrants to the UK. *Am J Hum Biol* [Internet]. 2013 [cited 2019 Sep 6];25:83–93. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=23175465](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=23175465)
- [39] Bar-Sadeh B, Amichai OE, Pnueli L, et al. Epigenetic regulation of 5 $\alpha$  reductase-1 underlies adaptive plasticity of reproductive function and pubertal timing. *BMC Biol* [Internet]. cited 2022 Jan 15;20:1–14. <https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-021-01219-6>.
- [40] 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 (Albany NY)* [Internet]. 2016 [cited 2022 Jan 24];8:1034. </pmc/articles/PMC4931852/>
- [41] Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* [Internet]. 2013 [cited 2017 May 11];14:R115. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=24138928](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=24138928)
- [42] Horvath S, Oshima J, Martin GM, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany NY)*. 2018;10:1758–1775.
- [43] Choi M, Genereux DP, Goodson J, et al. Epigenetic memory via concordant DNA methylation is inversely correlated to developmental potential of mammalian cells. *PLoS Genet* [Internet]. 2017;13:e1007060. <https://dx.plos.org/10.1371/journal.pgen.1007060>
- [44] Jones KB, Furukawa S, Marangoni P, et al. Quantitative clonal analysis and single-cell transcriptomics reveal division kinetics, hierarchy, and fate of oral epithelial progenitor cells. *Cell Stem Cell* [Internet]. 2019 [cited 2022 Sep 8];24:183–192.e8. <http://www.cell.com/article/S193459091830496X/fulltext>

- [45] Hannon E, Mansell G, Walker E, et al. Assessing the co-variability of DNA methylation across peripheral cells and tissues: implications for the interpretation of findings in epigenetic epidemiology. *PLoS Genet* [Internet]. 2021. [cited 2022 Jan 16];17:e1009443. <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009443>
- [46] Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* [Internet]. 2012[cited 2022 Sep 8]; 13:1–16. <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-86>
- [47] Liu Y, Aryee MJ, Padyukov L, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* [Internet]. 2013 [cited 2022 Sep 8];31:142–147. <https://pubmed.ncbi.nlm.nih.gov/23334450/>
- [48] Horvath S. New methylation age calculator. <https://dnamage.genetics.ucla.edu/new>
- [49] Levine ME, Lu AT, Chen BH, et al. Menopause accelerates biological aging. *Proc Natl Acad Sci U S A* [Internet]. 2016 [cited 2022 Jan 18];113:9327–9332. <https://www.pnas.org/content/113/33/9327>
- [50] Vaiserman A. Developmental Tuning of Epigenetic Clock. *Front Genet* [Internet]. 2018 [cited 2020 Nov 23]; 9:584. <https://www.frontiersin.org/article/10.3389/fgene.2018.00584/full>
- [51] Monseur B, Murugappan G, Bentley J, et al. Epigenetic clock measuring age acceleration via DNA methylation levels in blood is associated with decreased oocyte yield. *J Assist Reprod Genet* [Internet]. 2020 [cited 2021 Sep 3];37(5):1097–1103. <https://link.springer.com/article/10.1007/s10815-020-01763-0>
- [52] Richards JS, Ascoli M. Endocrine, paracrine, and autocrine signaling pathways that regulate ovulation. *Trends Endocrinol Metab* [Internet]. 2018 [cited 2021 Sep 3];29:313–325. <http://www.cell.com/article/S1043276018300432/fulltext>
- [53] Zhang Z, Wu L, Diao F, et al. Novel mutations in LHCGR (luteinizing hormone/choriogonadotropin receptor): expanding the spectrum of mutations responsible for human empty follicle syndrome. *J Assist Reprod Genet* [Internet]. 2020 [cited 2022 Jan 26];37:2861. <https://pubmed.ncbi.nlm.nih.gov/33764211/>
- [54] O'Connor CE, Bentley GR, Apostolidou S, et al. Differential methylation in PGR may explain varying progesterone levels in migrant Bangladeshi women. *Am J Hum Biol*. 2009;21:263.
- [55] Arand J, Spieler D, Karius T, et al. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet* [Internet]. 2012;8:e1002750. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=22761581](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22761581)
- [56] Giehr P, Walter J. Hairpin bisulfite sequencing: synchronous methylation analysis on complementary DNA strands of individual chromosomes. *Methods Mol Biol* [Internet]. 2018[cited 2022 Sep 16]; 1708:573–586. <https://pubmed.ncbi.nlm.nih.gov/29224164/>
- [57] Martin-Herranz DE, Aref-Eshghi E, Bonder MJ, et al. Screening for genes that accelerate the epigenetic aging clock in humans reveals a role for the H3K36 methyltransferase NSD1. *Genome Biol* [Internet]. 2019[cited 2022 Sep 11]; 20:1–19. <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1753-9>
- [58] Miner BE, Stöger RJ, Burden AF, et al. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. *Nucleic Acids Res*. 2004;32(17):e135.
- [59] Laird CD, Pleasant ND, Clark AD, et al. Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc Natl Acad Sci U S A* [Internet]. 2004;101:204–209. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=314163&tool=pmcentrez&rendertype=abstract>
- [60] Stöger R. Hairpin-Bisulfite PCR. *Methods Mol Biol*. 2021;2198:287–299.