

RESEARCH

Open Access



# Age-related changes in DNA methylation in a sample of elderly Brazilians

Hayley Welsh<sup>1\*</sup>, Caio M. P. F. Batalha<sup>2</sup>, Weili Li<sup>3</sup>, Nadja C. Souza-Pinto<sup>2</sup>, Yeda A. O. Duarte<sup>4,5</sup>, Michel S. Naslavsky<sup>6</sup> and Esteban J. Parra<sup>1</sup>

## Abstract

**Background** Age-related changes in DNA methylation (DNAm) play a critical role in regulating gene expression. However, most epigenome-wide association studies have predominantly focused on individuals of European descent. This study aims to characterize longitudinal changes in DNAm patterns in a cohort of elderly Brazilian participants.

**Methods** DNAm profiles were collected approximately nine years apart from 23 elderly Brazilian individuals using the Illumina Infinium MethylationEPIC BeadChip. Using mixed-effects models, we examined changes in DNAm patterns using both quantitative age and binary timepoint (e.g., baseline vs. follow-up) as predictors of interest to identify differentially methylated positions (DMPs). Significant DMPs were compared with a list of previously identified age-related DMPs. Differentially methylated regions (DMRs) were also identified using *DMRcate*. Gene ontology (GO) pathway enrichment analyses were performed to explore the functional significance of identified DMPs and DMRs.

**Results** Of the 586,229 autosomal probes included in the differential methylation analyses, 2768 significant ( $FDR < 0.05$ ) age-associated DMPs (aDMPs) and 2757 significant ( $FDR < 0.05$ ) timepoint-associated DMPs (tpDMPs) were identified. Of the 2768 aDMPs, 1471 were replicated from previous studies. Of the 1297 non-replicated CpGs, 77.4% were exclusive to the EPIC array. The DMR analyses identified 305 age-associated DMRs (aDMRs) and 372 time-point-associated DMRs (tpDMRs). Both aDMPs and aDMRs exhibited age-related hypermethylation within CpG islands and promoter regions of the genome, whereas hypomethylation predominantly occurred in interCGI and intergenic regions and introns. The GO enrichment analyses identified several neurological and cognition-related pathways enriched for hypermethylated CpG islands, many of which were mapped near transcription start sites and first exon regions.

**Conclusions** This longitudinal study identified age-associated and timepoint-associated DMPs and DMRs in a sample of elderly Brazilians. Most of the non-replicated CpGs were found to be on the new EPIC array, suggesting that more age-related studies using the EPIC array are required to validate these CpGs. The GO pathway enrichment analyses identified age-related enrichment of several gene sets related to cognitive and physical decline in elderly populations. The enrichment of these sites could provide evidence for age-related neurodegeneration and cognitive decline in elderly populations.

**Keywords** DNA methylation, Illumina EPIC array, Molecular aging, Longitudinal study

\*Correspondence:

Hayley Welsh  
hayley.welsh@mail.utoronto.ca

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Introduction

The process of aging can be understood as the progressive deterioration of biological system integrity arising from the accumulation of cellular level changes [1]. Variability in human aging and onset of age-related disease is not only based on genetics, but also epigenetics. DNA methylation (DNAm) is an important epigenetic mechanism that plays a role in gene expression and regulation [2, 3]. Methylation occurs when the enzyme DNA methyltransferase covalently bonds a methyl group ( $\text{CH}_3$ ) to the fifth carbon ring of a cytosine within a cytosine-guanine dinucleotide (CpG site), creating 5-methylcytosine (5mC) during synthesis [3–6]. Within the human genome there are approximately 28 million CpG sites, of which 60–90% are typically methylated [3, 5]. However, while most CpG sites are heavily methylated, CpG islands—areas of the genome containing higher densities of CpG sites—are typically resistant to methylation [2, 3, 5]. Dysregulation in DNAm is associated with premature aging and a variety of age-related illnesses including cancer, diabetes, cardiovascular disease, metabolic disease, and neurological diseases [5, 7]. Therefore, analyzing DNAm can provide important information regarding disease risk and health outcomes.

When it comes to age-related DNAm changes, studies have found a general pattern for DNAm over the lifespan, with methylation levels increasing in blood and saliva during the first year of life [8–11]. The DNAm changes observed in the first year of life primarily occur at intra-genic regions, CpG island shores, enhancers, and promoters that lack CpG islands [9, 12]. Throughout the developmental period of childhood and adolescence, DNAm levels increase rapidly [13–15]. DNAm changes then stabilize in early to middle adulthood [15]. From mid-adulthood to advancing age, overall DNAm levels decrease and interindividual variability increases [16]. This phenomenon is commonly referred to as “epigenetic drift,” characterized by non-directional changes in DNAm over time that involve both hypomethylation and hypermethylation events, resulting from both genetic and environmental factors [17, 18].

Age-associated hypomethylation, or demethylation, typically occurs in heterochromatic regions of the genome, such as repetitive elements and transposons, which contain the majority of CpG sites in the human genome [19]. Additionally, region- and site-specific hypermethylation has also been broadly observed in the aging genome [20]. Typically, age-related hypermethylation occurs at CpG islands, and more particularly CpG island promoters [21, 22]. Hypermethylation of CpG island promoters has been associated with the silencing of genes and suppression of transcription [21, 22]. Age-related changes in DNAm can be assessed through

examining differentially methylated positions (DMPs) and differentially methylated regions (DMRs) [23]. Biological relevance can be assigned to DMPs and DMRs using gene enrichment analysis [23].

When considering age-related changes in DNAm, comparatively few longitudinal studies have evaluated how DNAm profiles change over time in elderly populations [18] and ethnically diverse populations [24, 25]. In this study, we examined age-related changes in genome-wide DNAm in a sample of elderly Brazilian participants from the Health, Well-being and Aging (Saúde, Bem-estar e Envelhecimento; SABE) study cohort [26, 27]. Two samples taken ~9 years apart were available for each individual. Differential methylation analysis was performed using mixed-effects models with (1) quantitative age or (2) binary timepoint as predictors of interest. Positions and regions identified as being differentially methylated were then used for gene pathway enrichment analysis to identify differentially methylated pathways.

## Materials and methods

### Study participants and samples

The current study used whole blood samples obtained from the SABE study cohort, which is comprised of census-withdrawn elderly individuals from São Paulo, Brazil. This study cohort has been followed up every five years since 2000, with DNA first collected in 2010, and has been previously described in detail in recent genomic studies [27, 28]. Samples from 24 elderly adults (13 males and 11 females) were taken at two timepoints ( $9 \pm 0.71$  years apart) for a total of 48 samples. The first timepoint is from the 2010 collection wave, performed from June 2010 through September 2012, and the samples for the second timepoint were collected during the COVID-19 monitoring project, performed from June 2020 through April 2021. The 24 individuals were  $67.41 \pm 5.52$  years of age (mean  $\pm$  standard deviation) at timepoint one, and  $76.41 \pm 6.17$  at timepoint two. Global ancestry analyses in the SABE cohort have indicated evidence of genetic similarities with European, West African, Indigenous American, and East Asian populations, as well as broad variation in individual ancestral proportions, in agreement with the known demographic history of the Brazilian population [27].

### Blood collection and processing

Genomic DNA was extracted from whole peripheral blood samples collected in EDTA tubes. DNA extraction and purification followed manufacturer's recommended protocols, using the Qiagen AutoPure LS kit with Gentra automated extraction for samples collected at the first timepoint and manual extraction for samples collected at the second timepoint. The change in extraction

techniques was necessitated by the discontinuation of the equipment, however, the same commercial reagents were used for both timepoints. DNA was quantified with a Nanodrop spectrometer and diluted to 50 ng/uL. Whole-genome sequencing (WGS) data for the samples described above are also available.

#### **Characterization of DNA methylation using the EPIC array**

Approximately 1000 ng of human genomic DNA was used for bisulfite conversion. In August of 2021, DNA methylation was measured using the Infinium MethylationEPICv1.0 array at The Centre for Applied Genomics (TCAG, Hospital for Sick Children, Toronto, Ontario, Canada), in accordance with the protocols recommended by Illumina (San Diego, California, USA).

#### **Processing of DNA methylation DNA**

Using the R/Bioconductor packages *Meffil* (v1.1.0), *RnBeads* (v2.6.0), *minfi* (v1.34.0), and *wateRmelon* (v1.32.0), the methylation data were imported and processed, and quality control (QC) analyses were performed. First, we used *Meffil* to infer sex, which was then compared to reported sex. Using the 59 SNP probes that are available as part of the EPIC array, the concordance between the methylation intensities of the samples and the corresponding genotype calls extracted from their WGS data was calculated. Then, using the *RnBeads* QC pipeline, we performed comprehensive sample-level and probe-level QC. Specifically, probes were removed if their target sequences overlap with a SNP at any base, if they are known to be cross-reactive, and if more than 5% of the samples had a missing value. Additionally, the iterative GreedyCut algorithm was used to filter out samples and probes using a detection p-value threshold of 0.01. We used the *wateRmelon* package to extract bead numbers from the IDAT files and calculated the proportion of samples with a bead number < 3. Probes were removed if more than 5% of samples had a low bead number (< 3). We also computed detection p values using out-of-band probes empirical distribution with the *pOOBAH()* function in the *SeSAmE* (v1.14.2) R package, with a p value threshold of 0.05, and the *combine.neg* parameter set to TRUE. The *pOOBAH* was done in parallel with the previously mentioned QC steps, and the resulting probes flagged in both analyses were combined and removed from the data. These QC steps were previously described in Welsh et al. [29].

One sample was determined to be low quality based on QC, and due to the paired study design, we also excluded the second sample from the same individual. Therefore, a total of  $n=46$  samples from  $n=23$  individuals (13 males and 10 females) were included in the differential methylation analyses. Post QC, a total of 666,485 probes

remained, and we further excluded 15,119 sex chromosome probes. Since probes with low variation across all samples are less likely to show significant association, and to reduce the burden of multiple-testing correction [30], we decided to further exclude 65,137 probes for which the standard deviations of the beta values were in the lowest 10% of distribution. In summary, a total of 586,229 probes were analyzed.

#### **Differential cell composition estimation**

Different cell types may exhibit different methylation patterns, and differential cell type proportions across study samples can lead to spurious associations [31]. The proportions of cell types in each of the study samples were estimated using the method proposed by Salas et al. [32], as implemented in the R package *FlowSorted.Blood.EPIC*. Proportions of the 6 cell types, CD8+ T cell (CD8T), CD4+ T cell (CD4T), natural killer (NK), B cell (Bcell), monocytes (Mono) and neutrophils (Neu) were estimated for all 46 samples.

#### **Normalization and creation of cell composition adjusted M-values**

Post QC, we first performed data normalization using *SeSAmE* [33] as suggested in Welsh et al. [29]. We then computed adjusted M-values by removing the effect of differential cell type compositions through a regression-based approach, as described in Jones et al. [34]. Briefly, (1) *SeSAmE* normalized beta values were regressed on the estimated cell proportions for CD8T, CD4T, NK, Bcell, and Mono; estimated proportions for Neu were not included as a predictor due to multicollinearity, (2) the residuals from the linear regression models were extracted and the mean beta values were added to the residuals to obtain adjusted beta values, (3) the adjusted beta values were then converted to adjusted M-values. The adjusted M-values reflect sample methylation levels after removing the effect of differential cell type proportions.

#### **Differential methylation analyses**

Mixed-effects models were used for the differential methylation analysis. Two models were evaluated using the adjusted M-values modeled as the response variable, the first using quantitative age as the predictor of interest and the second using timepoint as a binary predictor of interest. For both mixed-effects models, a linear model was fitted with individual ID as a random effect using the *dream()* function in the R/Bioconductor package *variancePartition* (v1.33.11) [35], to account for the correlation between the two measurements from the same individual. Additionally, variables related to technical variation such as array

slide (Sentrix\_ID) and position (Sentrix\_Position) were also fitted as random effects. Since including individual ID as a random effect can account for subject-invariant variables such as sex and ancestry, they were not included as covariates. Similarly, smoking status was also found to be invariant between the two timepoints and was likewise not included as a covariate. Therefore, the final model for quantitative age was Adjusted M-values ~ Age + (1|Sentrix\_ID) + (1|Sentrix\_Position) + (1|Individual ID) and the final model for timepoint was Adjusted M-values ~ Binary Timepoint + (1|Sentrix\_ID) + (1|Sentrix\_Position) + (1|Individual ID). To account for multiple hypothesis testing, the Benjamini-Hochberg FDR corrected p-values were calculated, and probes with FDR-adjusted  $p\text{-value} < 0.05$  were considered statistically significant at the epigenome-wide level.

DMR analyses for each model were performed using the R package *DMRcate* (v2.16.1) [36]. Using the *rmSNPandCH()* function in *DMRcate*, an additional  $n = 5172$  probes were removed, leaving 581,057 probes for the DMR analyses. Because *DMRcate* uses *limma* for linear modeling, we altered the *cpg.annotate()* function code to replace the *limma lmFit()* function with the *variancePartition dream()* function. The CpGannotate object generated from the altered code was then processed using the *dmrcate()* function. A minimum number of three CpGs per region was used.

Genomic annotation was performed for both DMPs and DMRs using the R/Bioconductor package *annotatr* (v1.28.0). *Annotatr* provided CpG annotations (i.e., CpG island, shore, shelf, and interCGI), genic annotations (i.e., promoter, exon, intron, 1to5kb, 5UTR, 3UTR and intergenic), and gene annotations [37]. Shores are defined as 2 Kb upstream/downstream from the ends of the CpG island, shelves are defined as 2 Kb upstream/downstream from the farthest upstream/downstream limits of CpG shores, and interCGI regions, also known as open sea regions [37], are defined as the remaining genomic regions (i.e., > 4 Kb upstream/downstream from the ends of a CpG island). Some CpGs have multiple neighboring genes, meaning a single CpG might receive multiple genic annotations [38]. To mitigate this potential bias in the graphical representations of the data, fractional counting was used, where each genic annotation was divided by the total number of annotations for a given DMP or DMR (i.e., if a DMP or DMR mapped to a promoter region of gene A and gene B, each annotation was counted as 0.5) [38]. Chi-squared tests were then employed using the *chisq.test()* function, to determine if the observed CpG annotation distributions significantly ( $p < 0.05$ ) differ from expected values for both DMPs and DMRs.

### Gene expression/enrichment

Gene set enrichment analyses were conducted in order to identify potential pathways enriched in differentially methylated probes/regions. Using the R/Bioconductor package *missMethyl* (v1.36.0) gene ontology (GO) pathway enrichment analyses were performed. This package maps CpGs to genes and tests for gene enrichment using the Wallenius' non-central hypergeometric test, which accounts for gene-length bias [23, 39]. For each model (i.e., quantitative age and timepoint), gene enrichment analyses were performed using both the *gometh()* and *goregion()* functions which mapped the DMPs and DMRs to a list of 22,593 a priori genes, respectively. Enrichment analyses were performed on the total number of significant DMPs and DMRs as well as for all hyper- and hypomethylated DMPs and DMRs, hypermethylated DMPs and DMRs annotated to CpG islands, hypomethylated DMPs and DMRs annotated to interCGI (open sea) regions, and hypermethylated DMPs and DMRs annotated to CpG islands and mapped near transcription start sites (TSS200 and TSS1500) and 1st exon regions. Therefore, a total of 24 enrichment analyses were performed. All GO enrichment analyses were run using the default options of 'prior.prob = TRUE' and 'fract.counts = TRUE' to account for two additional sources of bias in the pathway analysis: (1) the differing number of probes per gene, and (2) CpGs could be annotated to multiple genes. The pathways for *gometh()* and *goregion()* include biological process (BP), cellular component (CC), and molecular function (MF) [39]. Pathways with an FDR-adjusted  $p\text{-value} < 0.05$  were considered statistically significant.

### Replication of previous studies

We compiled lists of significant age-associated DMPs (aDMPs) reported in previously published studies. The lists considered include the aDMPs of 33 studies collected by, and available through, the open platform EWAS Atlas [40] as well as lists from 15 additional studies [13, 41–54]. The EWAS Atlas did not include the exhaustive lists of aDMPs for six of the 33 studies; therefore, the full lists of these six of the studies were also considered [55–60]. These studies were conducted on a variety of tissue types, populations, age ranges, Illumina array types (i.e., 27 k, 450 k, and EPIC array), and study designs (e.g., cross-sectional, longitudinal, twin-study design). We compared these lists with our significant age- and timepoint-associated DMPs in order to determine the proportion of replicated CpGs.

### Results

In this study, the genome-wide methylation profiles of  $n = 46$  whole blood samples, from  $n = 23$  Brazilian participants, were examined. Two DNAm

profiles collected ~9 years apart from each individual were obtained using the Illumina Infinium MethylationEPIC BeadChip. After data preprocessing and *SeSAMe* normalization, 586,229 autosomal probes were eligible for analysis. Using the methylation status of these probes, we analyzed the effects of quantitative age and timepoint on DNA methylation profiles.

### Longitudinal EWAS on age

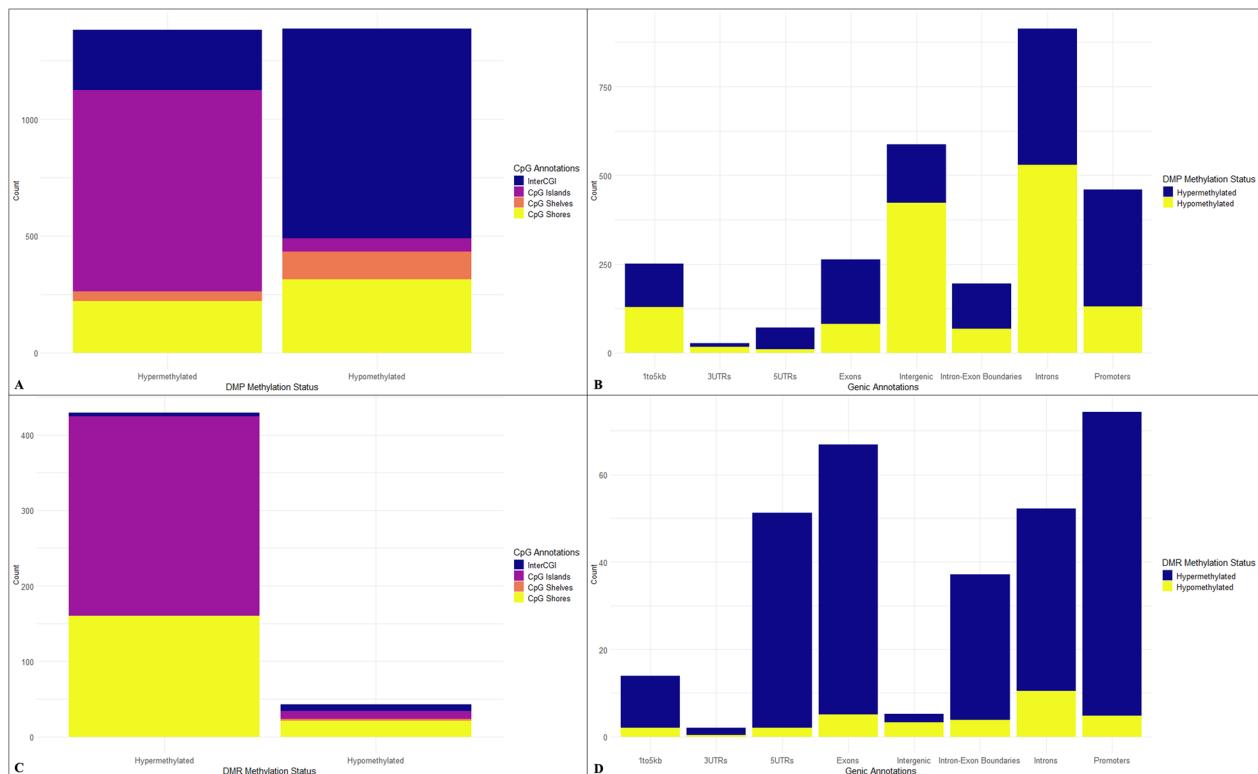
#### Differential methylation

Using a mixed-effects model with quantitative age as the predictor of interest, a total of 2768 age-associated significantly ( $FDR < 0.05$ ) differentially methylated positions (aDMPs) were identified. Using the more stringent Bonferroni correction ( $p < 0.05$ ), only  $n = 22$  significant probes were identified. The aDMPs identified account for 0.47% of investigated probes. The list of significant aDMPs can be found in File S1. We sought to determine the extent to which the significant aDMPs identified in the current study are replicated from previous research that considered age-related changes in DNA methylation. Comparing the significant aDMPs with age-associated probes identified by previous studies, 1471/2768 (53.1%) were replicated. Of the 1297 non-replicated CpGs, 1004

(77.4%) were exclusive to the EPIC array. The list of non-replicated aDMPs can be found in File S2.

Significant aDMPs with a  $\log FC > 0$  were considered hypermethylated (increasing with age) and significant aDMPs with a  $\log FC < 0$  were considered hypomethylated (decreasing with age). Nearly an equal number of hyper- and hypomethylated aDMPs were identified, with  $n = 1381$  hypermethylated probes and  $n = 1387$  hypomethylated probes. Differentially methylated regions (DMRs) were identified using *DMRcate* [36]. After filtering out an additional  $n = 5172$  probes using the *rmSNPandCH()* function, as recommended by *DMRcate*, a total of 581,057 EPIC array v1.0 autosomal probes were considered for the DMR analysis. *DMRcate* identified 2843 significant ( $FDR < 0.05$ ) aDMRs. Using the 2843 significant probes, 305 age-associated DMRs (aDMRs) were identified. Of the 305 aDMRs, 271 were found to be hypermethylated (mean difference  $> 0$ ) and 32 were found to be hypomethylated (mean difference  $< 0$ ). The list of aDMRs can be found in File S3.

Using CpG and genic annotations, the genomic region distributions of the hyper- and hypomethylated aDMPs and aDMRs were analyzed. Figure 1A and Table 1 show that hypermethylated aDMPs are primarily located



**Fig. 1** Distributions of CpG and genic annotations for significant aDMPs and aDMRs. **A** Distribution of CpG annotations for significant aDMPs, **B** Distribution of genic annotations for significant aDMPs using fractional counting, **C** Distribution of CpG annotations for significant aDMRs, **D** Distribution of genic annotations for significant aDMRs using fractional counting

within CpG islands, while hypomethylated aDMPs are primarily located within interCGI (open sea) regions. The chi-square results in Table 1 also show that more CpG island shelves and shores are hypomethylated than expected. Similarly, Fig. 1B and Table 2 show that there are proportionally more hypermethylated aDMPs in promoter, 5'UTR, exon, and intron-exon boundary regions, while hypomethylated aDMPs are primarily found in introns and intergenic regions. The aDMR results follow the same trend as the aDMP results, with hypermethylated aDMRs being primarily annotated to

CpG islands, and hypomethylated aDMRs being primarily annotated to interCGI regions (Fig. 1C; Table 3). The aDMR chi-square results in Table 3 similarly show that along with interCGI regions, there are more hypomethylated CpG island shelves and shores than expected. When considering genic annotations, there are more hypermethylated aDMRs than hypomethylated aDMRs in genic regions, but the opposite is observed in intergenic regions (Fig. 1D). With respect to the genic regions, when the hyper- and hypomethylated aDMRs are plotted separately (Fig. S1), it becomes clear that, compared

**Table 1** Chi-Square results for CpG annotations using significant aDMPs

|          | Hypermethylation (Observed) | Hypermethylation (Expected)* | Hypomethylation (Observed) | Hypomethylation (Expected)* |
|----------|-----------------------------|------------------------------|----------------------------|-----------------------------|
| InterCpG | 257                         | 576                          | 898                        | 579                         |
| Islands  | 861                         | 457                          | 55                         | 459                         |
| Shelves  | 43                          | 81                           | 119                        | 81                          |
| Shores   | 220                         | 267                          | 315                        | 268                         |
| Total    | 1381                        |                              | 1387                       |                             |

\*Rounded to the nearest whole number

X-squared = 1117.5, df = 3, p-value < 2.2e-16

**Table 2** Chi-Square results for genic annotations using significant aDMPs (fractional counting used)

|                      | Hypermethylation (Observed)* | Hypermethylation (Expected)* | Hypomethylation (Observed)* | Hypomethylation (Expected)* |
|----------------------|------------------------------|------------------------------|-----------------------------|-----------------------------|
| 1to5kb               | 122                          | 125                          | 129                         | 126                         |
| 3'UTR                | 10                           | 13                           | 17                          | 14                          |
| 5'UTR                | 61                           | 35                           | 9                           | 35                          |
| Exon                 | 183                          | 132                          | 81                          | 132                         |
| Intergenic           | 165                          | 293                          | 422                         | 294                         |
| Intronexonboundaries | 128                          | 98                           | 68                          | 98                          |
| Intron               | 383                          | 456                          | 530                         | 457                         |
| Promoter             | 329                          | 230                          | 131                         | 230                         |
| Total                | 1381                         |                              | 1387                        |                             |

\*Rounded to the nearest whole number

X-squared = 319.82, df = 7, p-value < 2.2e-16

**Table 3** Chi-Square results for CpG annotations using aDMRs

|          | Hypermethylation (Observed) | Hypermethylation (Expected)* | Hypomethylation (Observed) | Hypomethylation (Expected)* |
|----------|-----------------------------|------------------------------|----------------------------|-----------------------------|
| InterCpG | 5                           | 13                           | 9                          | 1                           |
| Islands  | 264                         | 249                          | 10                         | 25                          |
| Shelves  | 0                           | 3                            | 3                          | 0                           |
| Shores   | 160                         | 164                          | 21                         | 16                          |
| Total    | 429                         |                              | 43                         |                             |

\*Rounded to the nearest whole number

X-squared = 92.62, df = 3, p-value < 2.2e-16

to hypomethylated aDMRs, the relative proportions of hypermethylated aDMRs are higher within promoters, exons, and 5'UTR regions, than within intergenic regions. This result is further supported by the chi-square test results in Table 4 (using fractional counting) and Table S1 (without fractional counting/raw counts).

#### Gene enrichment analyses

To further investigate the functional significance of the aDMPs and aDMRs identified in this study, we performed Gene ontology (GO) enrichment pathway analyses. Gene enrichment analysis of aDMPs and aDMRs was conducted using 22,593 GO terms in the *missMethyl* package. Using all 2768 aDMPs, gometh() identified  $n=61$  significantly ( $FDR < 0.05$ ) enriched GO terms: 42 biological process (BP), 13 cellular component (CC), and six molecular function (MF). Using the topGSA() function, cell-cell signaling (GO:0007267), nervous system development (GO:0007399), and neurogenesis (GO:0022008) were identified as the top three most significant terms (Fig. S2A). When restricting the input to significant hypermethylated aDMPs,  $n=55$  significantly ( $FDR < 0.05$ ) enriched GO terms were identified: 33 BP, four CC, and 18 MF. The top three pathways for this analysis related to DNA-binding and RNA polymerase (GO:0000981, GO:0003700, and GO:0000977) (Fig. S2B). Using only hypermethylated aDMPs annotated to CpG islands,  $n=101$  significantly ( $FDR < 0.05$ ) enriched GO terms were identified: 55 BP, 18 CC, and 28 MF. The same top three pathways were identified as in the previous analysis (Fig. S2C), likely because, as shown in Table 1, the majority (62.3%) of age-associated hypermethylated CpGs are annotated to CpG islands. Restricting the gometh() analysis to hypermethylated aDMPs annotated to CpG islands and found near TSSs (i.e., TSS200 & TSS1500) and the 1st exon (i.e., promoter region), yielded

$n=51$  significant GO terms (Fig. S2D). No enriched GO terms were identified when using all hypomethylated aDMPs. However, two significantly ( $FDR < 0.05$ ) enriched co-occurring GO terms were identified when using hypomethylated open sea (interCGI) aDMPs: extracellular matrix (GO:0031012) and collagen-containing extracellular matrix (GO:0062023). All significant gometh() results can be found in File S4.

Using all  $n=305$  aDMRs, a total of  $n=22$  significantly ( $FDR < 0.05$ ) enriched GO terms were identified: 17 BP, four CC, and one MF (Fig. S3A). When only using the hypermethylated aDMRs,  $n=36$  significant ( $FDR < 0.05$ ) GO terms were identified: 25 BP, six CC, and five MF (Fig. S3B). Similarly, when only considering hypermethylated aDMRs annotated to CpG islands,  $n=33$  significant ( $FDR < 0.05$ ) GO terms were identified: 23 BP, five CC, and five MF (Fig. S3C). These three analyses identified the same top four pathways: nervous system development, neurogenesis, generation of neurons (GO:0048699), and synaptic membrane (GO:0097060). Restricting the gorenion() analysis to hypermethylated aDMRs annotated to CpG islands and found near TSSs (TSS200 and TSS1500) and 1st exon, yielded  $n=22$  significant GO terms (Fig. S3D). No pathways were identified when using hypomethylated aDMRs. All significant gorenion() results can be found in File S5.

#### Longitudinal EWAS on binary timepoint

##### Differential methylation

Using a mixed-effects model with binary timepoint as the predictor of interest, a total of 2757 timepoint-associated significantly ( $FDR < 0.05$ ) differentially methylated positions (tpDMPs) were identified. Using the more stringent Bonferroni correction ( $p < 0.05$ ), only  $n=9$  significant probes were identified. The tpDMPs identified account for 0.47% of investigated probes. The list of

**Table 4** Chi-Square results for genic annotations using aDMRs (fractional counting used)

|                      | Hypermethylation<br>(Observed)* | Hypermethylation<br>(Expected)* | Hypomethylation<br>(Observed)* | Hypomethylation<br>(Expected)* |
|----------------------|---------------------------------|---------------------------------|--------------------------------|--------------------------------|
| 1to5kb               | 12                              | 13                              | 2                              | 1                              |
| 3'UTR                | 2                               | 2                               | 0                              | 0                              |
| 5'UTR                | 49                              | 46                              | 2                              | 5                              |
| Exon                 | 62                              | 60                              | 5                              | 7                              |
| Intergenic           | 2                               | 4                               | 3                              | 1                              |
| Intronexonboundaries | 33                              | 33                              | 4                              | 4                              |
| Intron               | 42                              | 47                              | 10                             | 5                              |
| Promoter             | 69                              | 66                              | 5                              | 8                              |
| Total                | 271                             |                                 | 31                             |                                |

\*Rounded to the nearest whole number

X-squared = 22.239, df = 7, p-value = 0.00231

significant tpDMPs can be found in File S1. While using age and timepoint as a predictor of interest do not necessarily address the same questions, few EWAS looking at changes in DNA over time have used timepoint as a binary predictor [61, 62]. Therefore, as with the aDMPs, we sought to determine how many of the significant tpDMPs identified in the current study are replicated from previous research that considered age-related changes in DNA. Comparing the significant probes identified with age-associated probes identified by previous studies, 807/2757 (29.3%) of the tpDMPs were replicated. Of the 1950 non-replicated CpGs identified, 1334 (68.4%) were exclusive to the EPIC array. The list of non-replicated tpDMPs can be found in File S2.

Of the 2757 significant tpDMPs identified,  $n=778$  were found to be hypermethylated ( $\log FC > 0$ ) and  $n=1979$  were found to be hypomethylated ( $\log FC < 0$ ). *DMRcate* identified 2820 significant ( $FDR < 0.05$ ) tpDMPs. Using these significant probes, 372 timepoint-associated DMRs (tpDMRs) were identified, of which 60 were found to be hypermethylated (mean difference  $> 0$ ) and 312 were found to be hypomethylated (mean difference  $< 0$ ). The list of DMRs can be found in File S3.

Using CpG and genic annotations, the genomic region distributions of the hyper- and hypomethylated tpDMPs and tpDMRs were analyzed. In concordance with the aDMP results, Fig. S4A and Table S2 show that there are more hypomethylated tpDMPs than expected within interCGI (open sea) regions, whereas there are more hypermethylated tpDMPs than expected in CpG islands, and these differences are significant based on the chi-square test. When considering tpDMP genic annotations, all genomic categories are disproportionately more hypomethylated (Fig. S4B), but similarly to the analysis based on age, the number of hypomethylated probes in intergenic regions is higher than expected (Table S3). Though, unlike with the aDMP analysis, there is no disproportionately higher-than-expected hypermethylation occurring in promoter regions (Table S3). The tpDMR results for CpG annotations are also in line with the aDMR results, with more hypermethylated tpDMRs mapped to CpG islands and more hypomethylated tpDMRs mapped to interCGI regions than expected, based on the chi-square test results (Fig. S4C; Table S4). However, in contrast to the aDMR results, there are slightly more hypermethylated tpDMR in shores than expected (Table S4). As with the tpDMP genic annotations, there are more hypomethylated tpDMRs than hypermethylated tpDMRs for all the genomic categories, but similar to the analyses focused on age, the frequency of hypermethylated tpDMRs is higher in promoters, exons, and 1to5kb regions compared to intergenic regions (Fig. S5B). Additionally,

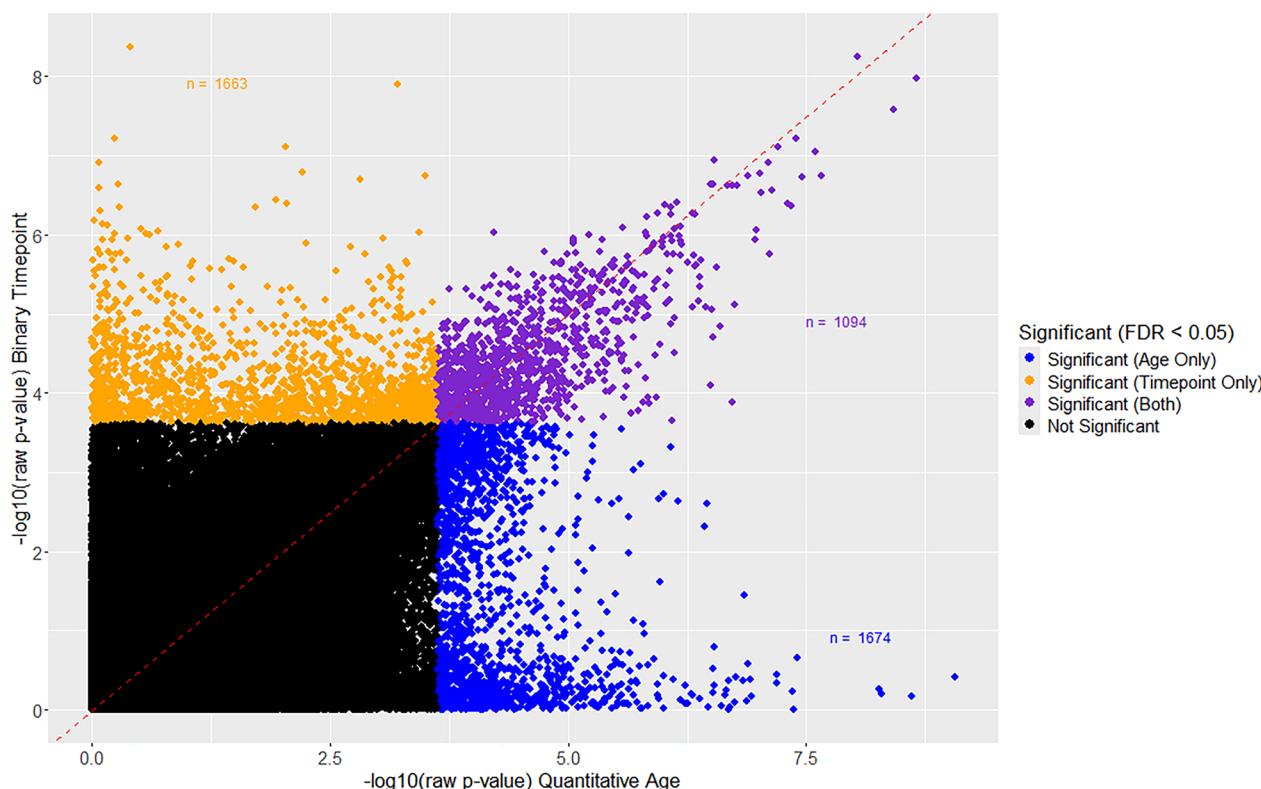
the chi-square results show that there are more hypermethylated tpDMRs annotated to promoter, 1to5kb, and 5'UTR regions than expected (Tables S5 & S6). Though it should be noted, that only the tpDMR chi-square results using the raw DMR genic annotation counts (Table S6), rather than fractional genic annotations counts (Table S5), are statistically significant.

#### Gene enrichment analyses

To further investigate the functional significance of the tpDMPs and tpDMRs identified in this study, we performed gene ontology (GO) enrichment pathway analyses. None of the gometh() analyses using the tpDMPs yielded significantly enriched pathways. When considering tpDMRs, no pathways were identified when considering all  $n=372$  tpDMRs or the  $n=60$  hypermethylated tpDMRs and  $n=312$  hypomethylated tpDMRs separately. However,  $n=39$  significant ( $FDR < 0.05$ ) GO terms were identified when restricting the analysis to hypermethylated tpDMRs annotated to CpG islands: 28 BP and 11 MF. The majority of the pathways were related to DNA-binding, transcription and RNA regulation, with the top three pathways being DNA-templated transcription (GO:0006351), regulation of DNA-templated transcription (GO:0006355), and RNA biosynthetic process (GO:0032774) (Fig. S6A). Restricting the goregion() analysis to hypermethylated tpDMPs annotated to CpG islands and found near TSSs (TSS200 and TSS1500) and 1st exon, yielded  $n=9$  significant GO terms (Fig. S6B). These terms included the previously mentioned top three pathways, as well as RNA metabolic process (GO:0016070), organic cyclic compound metabolic process (GO:1901360), organic cyclic compound biosynthetic process (GO:1901362), regulation of RNA metabolic process (GO:0051252), transcription regulator activity (GO: 0140110), and regulation of RNA biosynthetic process (GO:2001141). No pathways were identified when using hypomethylated tpDMRs mapped to interCGI regions. All significant goregion() results can be found in File S5.

#### Overlap between aDMPs and tpDMPs

When looking at the overlap of DMPs between the two models (i.e., aDMPs and tpDMPs),  $n=1094$  significant DMPs were identified (Fig. 2). All of the overlapping DMPs were found to be directionally concordant (i.e., hypermethylated or hypomethylated), with  $n=262$  hypermethylated probes and  $n=832$  hypomethylated probes. When considering the significant non-overlapping probes of each model, the majority (> 90%) of probes were directionally concordant between the two models.



**Fig. 2** Compares raw (unadjusted) p-values between aDMPs and tpDMPs. DMPs significant after FDR adjusted are highlighted ( $FDR < 0.05$ ). Orange represents significant DMPs only identified using timepoint as the predictor, blue represents significant DMPs only identified using quantitative age as the predictor, and purple represents the DMPs that overlap within both models

## Discussion

In this study, we explored longitudinal changes in DNAm patterns in a sample of elderly Brazilians. We considered both age- and timepoint-associated changes in DNAm. Overall, the results for both models are concordant in indicating a trend toward increased methylation with age within CpG islands, and decreased methylation within interCGI regions. Similarly, we observed increased methylation in promoters, exons and 5'UTR regions, when compared to introns and intergenic regions. The model based on timepoint tends to identify more hypomethylated probes and regions than the model based on age, but there is a substantial overlap between the probes identified in both models. When performing gene enrichment analyses with DMPs, only age-associated DMPs (aDMPs) provided significant results. Given that most comparable EWAS have considered quantitative age as opposed to binary timepoint, we will be primarily discussing the results of the age-associated analyses. Through this discussion we will explore the results of the age-associated analyses and contextualize them within the current research regarding age-related changes in DNAm.

Of the  $n=2768$  aDMPs identified, 53.1% were replicated from previous studies.  $N=551$  were replicated from the open platform EWAS Atlas [40] and an additional  $n=920$  were replicated from other previous studies (File S2) [13, 41–60]. Of the 1297 non-replicated aDMPs, 1004 (77.4%) are available on the EPIC array, but not in Illumina's previous methylation arrays. The lack of replication for these CpGs likely largely results from the fact that majority of previous studies use the older generation Illumina arrays (i.e., 27 k array and/or 450 k array), as only 7/33 studies included in the EWAS Atlas and only five out of the additional 15 study datasets considered probes available on the EPIC array. Additional factors that could have affected replication include differences in age ranges, study design (i.e., cross-sectional vs. longitudinal), tissue type, sample size, covariates considered, and population [63]. Aside from differences in arrays, age ranges are a particularly important factor to consider since changes in DNAm have been shown to follow different patterns throughout the life course [15]. Additionally, EWAS have predominantly been performed using populations of European descent [24, 25]. Therefore, additional longitudinal studies using the EPIC array and

ancestrally diverse populations may be required to validate the remaining 1297 non-replicated age-associated probes identified in this study.

### DNA methylation patterns

The association between aging and changes in DNAm has been studied for decades [20]. It was originally proposed that a spontaneous loss of 5-methylcytosines (5mCs) with age led to abnormal expression of genes in cells/tissues that are normally repressed [64]. Instead, a gradual loss of 5mC has been reported, with 5mC content being highest in early life and then decreasing gradually with age [65]. When considering changes that occur from mid-adulthood to advancing age, studies have shown overall DNAm levels decrease and interindividual variability increases [15, 16]. Additionally, region- and site-specific hypermethylation, typically occurring at CpG islands, has also been broadly observed in the aging genome [20]. As expected, the patterns of age-associated changes observed in the current study broadly follow the patterns of epigenetic drift typical for individuals of advancing age (elderly individuals). Hypermethylated aDMPs and aDMRs were found to be enriched in CpG islands and promoter regions (See Fig. 1). In contrast, clear patterns of hypomethylation were observed for interCGI (open sea) and intergenic regions.

### Enrichment analyses

Many pathways were enriched for hypermethylated probes and regions, especially those mapped to CpG islands. Several of the significantly enriched GO terms identified in this study have also been identified by other studies looking at age-associated changes in DNAm. For example, cell–cell signaling [53, 56, 66, 67], nervous system development [53, 55, 68], and neurogenesis [56] have been widely identified in other aging studies. Marttila et al. [50] also reported many of the same transcription and RNA regulatory GO terms identified in the current study.

Along with neurogenesis and generation of neurons, many other neurological/cognitive related GO terms were identified in the current study, such as neuron differentiation (GO:0030182), central nervous system (GO:0007417), brain development (GO:0007420), learning or memory (GO:0007611), and memory (GO:0007613). All of these processes are critically involved in aging and neurodegeneration [69] and have been identified in DNAm studies examining differential methylation patterns related to Alzheimer's disease [70]. Additionally, the majority of the hypermethylated aDMPs and aDMRs that were annotated to CpG islands were also mapped near TSSs (i.e., TSS200 and TSS1500) and

first exon regions (See File S3 & S4). Increased methylation near TSSs and first exon regions is often associated with transcriptional silencing [71] and gene expression suppression [20, 72]. Therefore, these GO terms may provide evidence for the accumulation of epigenetic changes relating to cognitive decline in elderly individuals.

When considering hypomethylated positions and regions, only age-related hypomethylated interCGI probes identified enriched GO pathways. These pathways were related to extracellular matrix and collagen functioning. Enrichment in extracellular matrices has been found to be associated with aging and a variety of age-related diseases, such as Pulmonary fibrosis [73], osteoarthritis [74, 75], and intervertebral disk degeneration [76]. Therefore, the results of this study suggest that the accumulation of age-related changes in DNAm, which primarily involved hypermethylation at CpG islands and hypomethylation at interCGI regions, may be affecting the regulation of gene expression in important regulatory regions of the genome.

### Conclusion

In conclusion, this study explored age-associated changes in DNA methylation in a sample of elderly Brazilians. The inclusion of longitudinal data and the focus on a population underrepresented in biomedical studies are two of the main strengths of the study. The major limitation is the small sample size, but despite this limitation, the majority of our age-associated probes overlap with the findings of previous EWAS of age. Most of the non-replicated CpGs were found to be on the new EPIC array, suggesting that more age-related studies using the EPIC array are required to validate these CpGs. Using gene ontology pathway enrichment analyses, we identified age-related enrichment of several gene sets, many of which are related to developmental and transcriptional regulation. The enrichment of these sites could be directly associated with age-related cognitive and physical decline in elderly populations, though larger longitudinal studies are required to properly investigate the relationship between these epigenetic changes and gene expression.

### Abbreviations

|          |   |
|----------|---|
| aDMP     | Age-associated differentially methylated position |
| aDMR     | Age-associated differentially methylated region   |
| BP       | Biological process                                |
| CC       | Cellular component                                |
| CpG      | Cytosine-guanine dinucleotide                     |
| DMP      | Differentially methylated position                |
| DMR      | Differentially methylated region                  |
| DNAm     | DNA methylation                                   |
| EWAS     | Epigenome-wide association study                  |
| FDR      | False discovery rate                              |
| GO       | Gene ontology                                     |
| interCGI | Areas of the genome > 4 Kb from a CpG island      |
| Kb       | Kilobases   |

|        |   |
|--------|---|
| logFC  | Log fold change   |
| MF     | Molecular function                                      |
| QC     | Quality control   |
| TCGA   | The Centre for Applied Genomics                         |
| SABE   | Saúde, Bem-estar e Envelhecimento                       |
| tpDMP  | Timepoint-associated differentially methylated position |
| tpDMR  | Timepoint-associated differentially methylated region   |
| TSS    | Transcription start site                                |
| WGS    | Whole-genome sequencing                                 |
| 1–5 kb | 1–5 Kilobases   |
| 3'UTR  | 3' Untranslated region                                  |
| 5mC    | 5-Methylcytosine  |
| 5'UTR  | 5' Untranslated region                                  |

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-025-01821-3>.

- Additional file1.
- Additional file2.
- Additional file3.
- Additional file4.
- Additional file5.
- supplementary Figures.
- supplementary Tables.

## Acknowledgements

The authors acknowledge the SABE participants for their long-term contributions as well as thank the Centre for Applied Genomics, particularly the Microarray Facility, at the Hospital for Sick Children, for processing the samples used in this study.

## Author contributions

EJP and MSN were responsible for data and funding acquisition, as well as project administration. CMPFB and WL performed quality control and data normalization. HW carried out formal analysis and visualization. HW prepared the original draft, which was reviewed and edited by EJP, CMPFB, and WL. YAOD was responsible for data curation. EJP, MSN, and NCSP also performed supervisory roles. All authors approved the final manuscript.

## Funding

This project was funded by a McLaughlin Centre Accelerator Award (MC-2020-11). Funding for M. Naslavsky was provided by grants FAPESP (SABE 2014/50649-6, INCT 2014/50931-3), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq INCT 465355/2014-5). Additionally, funding for H. Welsh was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Open access funding provided by the University of Toronto.

## Availability of data and materials

Due to privacy concerns, individual-level data cannot be made publicly available. However, data requests, describing the proposed use(s) of the data, can be sent to Hayley Welsh, the corresponding author for the study, or Dr. Esteban Parra (esteban.parra@utoronto.ca).

## Declarations

### Ethics approval and consent to participate

The SABE study was approved by the IRB from the Public Health School at the University of São Paulo (FSP-USP), protocol 2044, approval CAAE 47683115.4.0000.5421. All individuals enrolled in the SABE cohort provided written informed consent and the ethic protocols were approved by local and national institutional review boards COEP/FSP/USP OF COEP/23/10, CONEP 2044/2014, CEP HIAE 1263-10, University of Toronto RIS 39685.

## Consent for publication

Not applicable.

## Competing interests

The authors declare no competing interests.

## Author details

<sup>1</sup>Department of Anthropology, University of Toronto at Mississauga, Mississauga, Canada. <sup>2</sup>Department of Biochemistry, University of São Paulo, São Paulo, Brazil. <sup>3</sup>The Centre for Applied Genomics, Hospital for Sick Children, Toronto, Canada. <sup>4</sup>Medical-Surgical Nursing Department, School of Nursing, University of São Paulo, São Paulo, Brazil. <sup>5</sup>Epidemiology Department, Public Health School, University of São Paulo, São Paulo, Brazil. <sup>6</sup>Department of Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Brazil.

Received: 2 September 2024 Accepted: 17 January 2025

Published online: 05 February 2025

## References

- Belsky DW, Caspi A, Arseneault L, et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *Elife*. 2020;9: e54870.
- Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropharmacol*. 2012;38(1):23–38.
- Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet*. 2013;14(3):204–20.
- Al-Hasani K, Mathiyalagan P, El-Osta A. Epigenetics, cardiovascular disease, and cellular reprogramming. *J Mol Cell Cardiol*. 2019;128:129–33.
- Johnson AA, Akman K, Calimport SR, et al. The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation Res*. 2012;15(5):483–94.
- Mulligan CJ. Early environments, stress, and the epigenetics of human health. *Annu Rev Anthropol*. 2016;45:233–49.
- Brunet A, Berger SL. Epigenetics of aging and aging-related disease. *J Gerontol A Biol Sci Med Sci*. 2014;69(S1):S17–20.
- Herbstman JB, Wang S, Perera FP, et al. Predictors and consequences of global DNA methylation in cord blood and at three years. *PLoS ONE*. 2013;8(9): e72824.
- Martino DJ, Luke YJ, Gordon L, et al. Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol*. 2013;14(R42):1–14.
- Martino DJ, Tulic MK, Gordon L, et al. Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics*. 2011;6(9):1085–94.
- Wikenius E, Moe V, Smith L, et al. DNA methylation changes in infants between 6 and 52 weeks. *Sci Rep*. 2019;9:17587.
- McClay JL, Aberg KA, Clark SL, et al. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Hum Mol Genet*. 2014;23(5):1175–85.
- Alisch RS, Barwick BG, Chopra P, et al. Age-associated DNA methylation in pediatric populations. *Genome Res*. 2012;22(4):623–32.
- Lister R, Mukamel EA, Nery JR, et al. Global epigenomic reconfiguration during mammalian brain development. *Science*. 2013;341(6146):1237905.
- Tsang SY, Ahmad T, Mat FW, et al. Variation of global DNA methylation levels with age and in autistic children. *Hum Genomics*. 2016;10(31):1–6.
- Talens RP, Christensen K, Putter H, et al. Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell*. 2012;11(4):694–703.
- Boks MP, Derkx EM, Weisenberger DJ, et al. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS ONE*. 2009;4(8): e6767.
- Tan Q, Heijmans BT, Hjelmborg JV, et al. Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort. *Int J Epidemiol*. 2016;45(4):1146–58.

19. Ciccarone F, Tagliatesta S, Caiafa P, et al. DNA methylation dynamics in aging: How far are we from understanding the mechanisms? *Mech Ageing Dev.* 2017;174:3–17.
20. Xiao FH, Wang HT, Kong QP. Dynamic DNA methylation during aging: a “prophet” of age-related outcomes. *Front Genet.* 2019;10(107):1–8.
21. Reale A, Tagliatesta S, Zardo G, et al. Counteracting aged DNA methylation states to combat ageing and age-related diseases. *Mech Ageing Dev.* 2022;206(111695):1–20.
22. Zampieri M, Ciccarone F, Calabrese R, et al. Reconfiguration of DNA methylation in aging. *Mech Ageing Dev.* 2015;151:60–70.
23. Campagna MP, Xavier A, Lechner-Scott J, et al. Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clin Epigenetics.* 2021;13(214):1–24.
24. Breeze CE, Beck S, Berndt SI, et al. The missing diversity in human epigenomic studies. *Nat Genet.* 2022;54(6):737–9.
25. Breeze CE, Wong JY, Beck S, et al. Diversity in EWAS: current state, challenges, and solutions. *Genome Med.* 2022;14(71):1–4.
26. Lebrão ML, Duarte YA, Santos JL, et al. 10 Years of SABE Study: background, methodology and organization of the study. *Rev Bras Epidemiol.* 2019;21: e180002.
27. Naslavsky MS, Sciliar MO, Yamamoto GL, et al. Whole-genome sequencing of 1,171 elderly admixed individuals from Brazil. *Nat Commun.* 2022;13(1004):1–11.
28. Naslavsky MS, Yamamoto GL, de Almeida TF, et al. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. *Hum Mutat.* 2017;38(7):751–63.
29. Welsh H, Batalha CM, Li W, et al. A systematic evaluation of normalization methods and probe replicability using Infinium EPIC methylation data. *Clin Epigenet.* 2023;15(41):1–12.
30. Meng H, Joyce AR, Adkins DE, et al. A statistical method for excluding non-variable CpG sites in high-throughput DNA methylation profiling. *BMC Bioinform.* 2010;11:1–10.
31. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 2014;15(R31):1–9.
32. Salas LA, Koestler DC, Butler RA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol.* 2018;19(64):1–14.
33. Zhou W, Triche TJ Jr, Laird PW, et al. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res.* 2018;46(20): e123.
34. Jones MJ, Islam SA, Edgar RD, et al. Adjusting for cell type composition in DNA methylation data using a regression-based approach. In: Haggarty P, Harrison K, editors., et al. Population epigenetics: methods and protocols. New York: Humana Press; 2015. p. 99–106.
35. Hoffman GE, Rousos P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinform.* 2021;37(2):192–201.
36. Peters TJ, Buckley MJ, Chen Y, et al. Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate. *Nucleic Acids Res.* 2021;49(19): e109.
37. Cavalcante RG, Sartor MA. Annotatr: genomic regions in context. *Bioinform.* 2017;33(15):2381–3.
38. Komaki S, Ohmomo H, Hachiya T, et al. Longitudinal DNA methylation dynamics as a practical indicator in clinical epigenetics. *Clin Epigenet.* 2021;13(219):1–12.
39. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. *Genome Biol.* 2021;22(173):1–26.
40. Xiong Z, Yang F, Li M, et al. EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. *Nucleic Acids Res.* 2022;50(D1):D1004–9.
41. Bocklandt S, Lin W, Sehl ME, et al. Epigenetic predictor of age. *PLoS ONE.* 2011;6(6): e14821.
42. Cruickshank MN, Oshlack A, Theda C, et al. Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. *Genome Med.* 2013;5(96):1–12.
43. Fernández AF, Bayón GF, Urdinguio RG, et al. H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. *Genome Res.* 2015;25:27–40.
44. Gim JA. Integrative approaches of DNA methylation patterns according to age, sex and longitudinal changes. *Curr Genomics.* 2023;23(6):385–99.
45. Heyn H, Li N, Ferreira HJ, et al. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci.* 2012;109(26):10522–7.
46. Higham J, Kerr L, Zhang Q, et al. Local CpG density affects the trajectory and variance of age-associated DNA methylation changes. *Genome Biol.* 2022;23(216):1–28.
47. Johansson Å, Enroth S, Gyllensten U. Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS ONE.* 2013;8(6): e67378.
48. Kananen L, Marttila S, Nevalainen T, et al. Aging-associated DNA methylation changes in middle-aged individuals: the Young Finns study. *BMC Genomics.* 2016;17(103):1–12.
49. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY).* 2018;10(4):573–91.
50. Marttila S, Kananen L, Häyrynen S, et al. Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics.* 2015;16(179):1–17.
51. Refn MR, Andersen MM, Kampmann ML, et al. Longitudinal changes and variation in human DNA methylation analysed with the Illumina MethylationEPIC BeadChip assay and their implications on forensic age prediction. *Sci Rep.* 2023;13(21658):1–10.
52. Steegenga WT, Boekschoten MV, Lute C, et al. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age.* 2014;36:1523–40.
53. Tajuddin SM, Hernandez DG, Chen BH, et al. Novel age-associated DNA methylation changes and epigenetic age acceleration in middle-aged African Americans and whites. *Clin Epigenet.* 2019;11(119):1–16.
54. Vershinina O, Bacalini MG, Zaikin A, et al. Disentangling age-dependent DNA methylation: deterministic, stochastic, and nonlinear. *Sci Rep.* 2021;11(9201):1–12.
55. Alsaleh H, Hadrill PR. Identifying blood-specific age-related DNA methylation markers on the Illumina MethylationEPIC® BeadChip. *For Sci Int.* 2019;303(109944):1–11.
56. Bell JT, Tsai PC, Yang TP, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy aging population. *PLoS Genet.* 2012;8(4): e1002629.
57. Rakyan VK, Down TA, Maslau S, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.* 2010;20:434–9.
58. Teschendorff AE, Menon U, Gentry-Maharaj A, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* 2010;20:440–6.
59. Xu Z, Taylor JA. Genome-wide age-related DNA methylation changes in blood and other tissues relate to histone modification, expression and cancer. *Carcinog.* 2013;35(2):356–64.
60. Yusipov I, Bacalini MG, Kalyakulina A, et al. Age-related DNA methylation changes are sex-specific: a comprehensive assessment. *Aging (Albany NY).* 2020;12(23):24057–80.
61. Urdinguio RG, Torró MI, Bayón GF, et al. Longitudinal study of DNA methylation during the first 5 years of life. *J Transl Med.* 2016;14(160):1–12.
62. Wang D, Liu X, Zhou Y, et al. Individual variation and longitudinal pattern of genome-wide DNA methylation from birth to the first two years of life. *Epigenetics.* 2012;7(6):594–605.
63. Gopalan S, Carja O, Fagny M, et al. Trends in DNA methylation with age replicate across diverse human populations. *Genet.* 2017;206(3):1659–74.
64. Zs-Nagy IM, Cutler RG, Semsei I. Dysdifferentiation hypothesis of aging and cancer: a comparison with the membrane hypothesis of aging. *Ann NY Acad Sci.* 1988;521:215–25.
65. Goel N, Karir P, Garg VK. Role of DNA methylation in human age prediction. *Mech Ageing Dev.* 2017;166:33–41.
66. Florath I, Butterbach K, Müller H, et al. 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.* 2014;23(5):1186–201.
67. Kim S, Wyckoff J, Morris AT, et al. DNA methylation associated with healthy aging of elderly twins. *GeroScience.* 2018;40:469–84.
68. Wang Y, Karlsson R, Lampå E, et al. Epigenetic influences on aging: a longitudinal genome-wide methylation study in old Swedish twins. *Epigenetics.* 2018;13(9):975–87.
69. Cosín-Tomás M, Álvarez-López MJ, Companys-Alemany J, et al. Temporal integrative analysis of mRNA and microRNAs expression profiles and

- epigenetic alterations in female SAMP8, a model of age-related cognitive decline. *Front Genet.* 2018;9:596.
- 70. Li QS, Sun Y, Wang T. Epigenome-wide association study of Alzheimer's disease replicates 22 differentially methylated positions and 30 differentially methylated regions. *Clin Epigenet.* 2020;12(149):1–14.
  - 71. Brenet F, Moh M, Funk P, et al. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS ONE.* 2011;6(1): e14524.
  - 72. Jiao Y, Widschwendter M, Teschendorff AE. A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. *Bioinform.* 2014;30(16):2360–6.
  - 73. Lu Y, Chen J, Wang S, et al. Identification of genetic signature associated with aging in pulmonary fibrosis. *Front Med.* 2021;8:744239.
  - 74. Shi S, Wan F, Zhou Z, et al. Identification of key regulators responsible for dysregulated networks in osteoarthritis by large-scale expression analysis. *J Orthop Surg Res.* 2021;16(259):1–13.
  - 75. Wang T, Li Z, Zhao S, et al. Characterizing hedgehog pathway features in senescence associated osteoarthritis through Integrative multi-omics and machine learning analysis. *Front Genet.* 2024;15(1255455):1–25.
  - 76. Hu S, Fu Y, Yan B, et al. Analysis of key genes and pathways associated with the pathogenesis of intervertebral disc degeneration. *J Orthop Surg Res.* 2020;15(371):1–11.

## Publisher's Note

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