nature communications

Article

https://doi.org/10.1038/s41467-026-76240-w

Nanopore sequencing identifies parent-of-origin specific age-associated methylation changes at imprinted loci in the human genome

Received: 2 December 2025

Accepted: 24 July 2026

Published online: 29 July 2026

Check for updates

Brynja Sigurpalsdottir[1,2] ✉, Guillaume Holley[1], Sverrir Þ. Sverrisson[1], Droplaug N. Magnusdottir[1], Pall I. Olafsson[1], Arnaldur Gylfason[1], Olafur Þ. Magnusson[1], Gisli Masson[1], Olafur A. Stefansson[1] & Bjarni V. Halldorsson[1,2] ✉

Aging is accompanied by widespread DNA methylation changes, yet their full genomic scope and parent-of-origin dynamics remain poorly understood. Here, we apply nanopore long-read sequencing to 7,284 whole blood samples enabling methylation measurements of 17,959,684 high-quality CpG units. Over 20% of the measured high quality CpG units undergo age-associated changes, predominantly hypomethylation. From these data, we construct a methylation aging clock from 1,373 high-quality CpG units, with median absolute prediction error of 2.43 years. Importantly, phasing the methylation to parental haplotypes enables systematic analysis of age effects in parent-of-origin specific context, uncovering 702 high-quality CpG units with parent-of-origin specific age-association, most of which are located at imprinted genomic regions. At the *DIRAS3* locus, we detect age-dependent hypermethylation on the active paternal allele, indicative of attenuation of parent-of-origin specific methylation with age. Together, these findings establish nanopore sequencing as a powerful tool for mapping both genome-wide and parent-of-origin specific signatures of methylation aging and provide evidence that methylation patterns at imprinted loci become progressively altered with age.

DNA methylation, primarily occurring at cytosines within CG dinucleotides (CpG sites), has long been suspected to have roles in maintaining cellular identity and cell-type specification[1], X chromosome inactivation[2] and chromatin and transcriptional regulation[3]. As individuals age, changes in methylation levels occurs[4]. These methylation changes form the basis for accurate methylation clocks that aim to estimate biological age[5–8]. While these clocks have enabled valuable insights into age-related phenotypes and risk stratification[9,10], their mechanistic foundation remain poorly understood[11].

A key limitation of previous studies[3–11] lies in their reliance on array-based technologies, which cannot resolve methylation at the level of individual haplotypes. Additionally array-based methods only interrogate fraction of ~30 million CpG sites found in the human genome. As a result, little is known about how aging affects allele-specific methylation patterns or the genome-wide changes with age. This is especially relevant for imprinted genomic regions, where gene expression is parent-of-origin specific and is frequently regulated by allele-specific methylation[11,12]. While methylation within these regions

[1]Amgen deCODE genetics, Sturlugata 8, Reykjavik, Iceland. [2]School of Technology, Reykjavik University, Menntavegur 1, Reykjavik, Iceland.
✉e-mail: brynjas@decode.is; bjarnih@decode.is

has traditionally been considered stable[13], recent studies suggest that genomic imprinting fidelity may deteriorate at imprinted loci with age[10]. These studies have reported age-associated changes in methylation and imprinting stability at selected loci, particularly in brain tissue and in analysis based on array technologies[10]. However, the genome-wide extent of parent-of-origin specific methylation changes with age at imprinted genomic loci remains to be fully characterized.

Results

To map the landscape of age-associated changes in methylation, we analyzed whole blood samples from 7284 Icelanders (3497 males, 3787 females), using Oxford nanopore long-read sequencing (LRS)[14,15]. CpGs were measured as units if they were located within 10 base pairs (bp) of each other, and consequently, we refer to them as CpG units[14]. Here, 17,959,684 autosomal high-quality-CpG units (hq-CpGs), including 22,909,253 CpG sites, satisfied our criteria for high-quality methylation measurements (Supplementary Data 1)[14].

3,739,875 hq-CpGs (20.8%, Fig. 1A) were significantly associated with chronological age after covariate adjustment, including principal components which partially represent blood cell composition (Supplementary Fig. 1), and Bonferroni correction ($p < 2.8 \cdot 10^{-9} = 0.05/18 \cdot 10^6$, Supplementary Data 2). Across 10 permutations, only one hq-CpG reached Bonferroni significance in a single round, indicating that genome-wide significant findings are unlikely to arise due to random variation (Supplementary Table 1). The mean absolute effect size is 0.0005, indicating that age-associated hq-CpGs change their methylation ratio by an average of 0.0005 per year of age. We found that the mean methylation of age-associated hq-CpGs ($\mu = 0.74$) is lower than that of hq-CpGs without age association ($\mu = 0.80$, Welch two sample t-test, $p < 2.2 \times 10^{-16}$, Supplementary Fig. 2A). The vast majority, or 92.4%, exhibited progressive hypomethylation (Linear regression, $p < 2.8 \cdot 10^{-9}$, $\beta < 0$; negative effect size), whereas 7.6% showed hypermethylation (Linear regression, $p < 2.8 \cdot 10^{-9}$, $\beta > 0$; positive effect size) with age (Fig. 1A). Age-associated hypomethylation mostly occurs in highly methylated hq-CpGs, while age-associate hypermethylation spanned a wider range of methylation levels (Supplementary Fig. 2B). These results are consistent with loss of methylation fidelity with age (Fig. 1B)[4]. In a subset of 1,934 samples with available blood cell counts, 83.9% of hq-CpGs remained nominally significantly associated with age after adjustment for cell counts ($p < 0.05$, $\text{sign}(\beta_1) = \text{sign}(\beta_2)$). This suggests that age-associated methylation changes are not solely driven by age-related shifts in blood cell composition (Supplementary Table 2). Age-associated hq-CpGs showed distinct patterns of enrichment across genomic regions depending on whether they exhibited gain or loss of methylation (Fig. 1C). Consistent with previous studies[5,16,17], hq-CpGs that showed hypermethylation with age were enriched in repressed polycomb regions (ReprPC), flanking TSS, bivalent enhancers, CpG islands, and coding exons (Fig. 1C, upper panel). Most of the age-associated hq-CpGs however, showed hypomethylation with age, and these hq-CpGs were more prevalent in introns, intergenic and quiescent regions (Fig. 1C, lower panel). We further demonstrated that age-associated hq-CpGs exhibit higher residual variance compared to non-age-associated hq-CpGs, indicating more unexplained variance in their methylation levels (Supplementary Fig. 3). Taken together, these patterns suggest that most of the observed age-associated methylation changes occur within non-functional, less stable genomic regions[6,18].

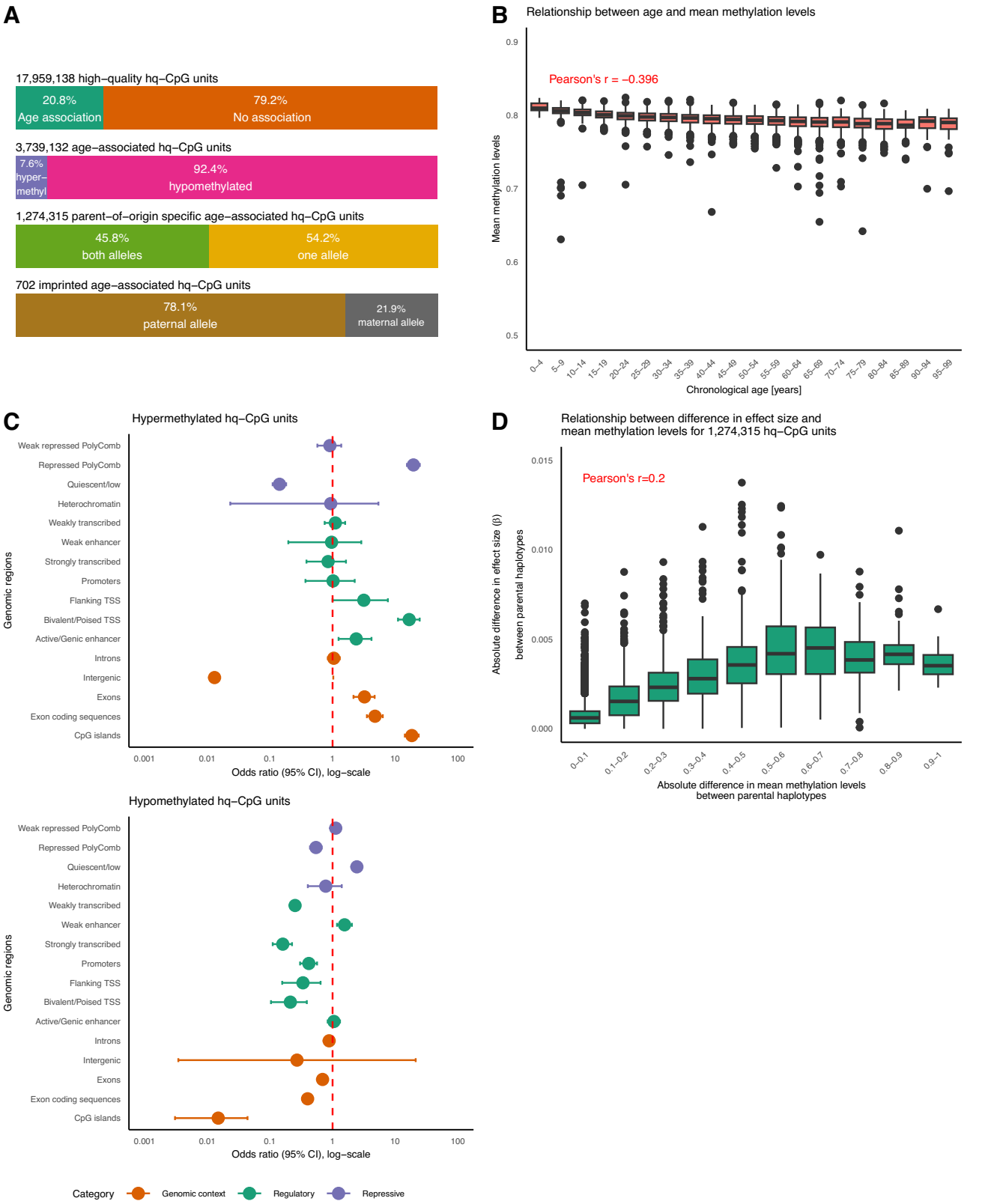
We developed a methylation clock based on 1,373 hq-CpGs (Supplementary Data 3), achieving median absolute error (medAE) of 2.22 years and 2.43 years, in training and test set, respectively. Our clock achieves similar prediction errors as two gold standard methylation clocks: Horvath's multi-tissue methylation clock (train medAE = 2.9 years, test medAE = 3.6 years, nCpGs = 353)[5] and Hannum's methylation clock (overall medAE = 3.9 years, test medAE = 4.9 years, nCpGs = 71)[6]. Both Horvath's and Hannum's clocks have been extensively tested across a wide range of tissues, populations, and cohorts, whereas the generalizability of our clock remains to be established through broader external validation. Model performance remained robust even after hq-CpGs identified as predictive in the model, i.e. those with non-zero coefficient, were iteratively removed and the model retrained on the remaining hq-CpGs, indicating redundancy among predictive sites (Supplementary Note 1, Supplementary Table 3). 398 (29.0%) and 975 (71.0%) of the hq-CpGs included in the clock were hyper- and hypomethylated, respectively. This distribution of hyper- and hypomethylated hq-CpGs contrast sharply with that observed in the age-associated CpG set, where hypermethylated hq-CpGs where 4.95 times more frequent (chi-square test, $p < 1.37 \cdot 10^{-195}$).

1,274,315 hq-CpGs (7.1%) showed age-associated methylation changes on one (54.2%) or both (45.8%) of the two parental haplotypes after Bonferroni correction ($p < 2.8 \cdot 10^{-9}$, Fig. 1A, Supplementary Data 4). Of these hq-CpGs, 99.7% were previously associated with age without consideration of parent-of-origin effects, indicating that age associations are largely shared between parental alleles. Effect sizes across haplotypes were highly correlated (Pearson's $r = 0.94$, Supplementary Fig. 4), reflecting symmetric aging effects at most loci. Interestingly, 702 hq-CpGs across the genome (Supplementary Data 5), showed a significant difference in effect size (z-test, $p < 3.9 \cdot 10^{-8} = 0.05/1{,}274{,}315$). Permutation testing indicated that approximately 0.3% of hq-CpGs reached this significance threshold under random age assignments (Supplementary Table 4). 300 hq-CpGs (42.7%) were not amongst the ~3.7 million age-associated hq-CpGs, identified using unphased data, suggesting that these age effects cannot be identified without using parent-of-origin assigned methylation. Differences in effect sizes were correlated with differences in methylation levels between the two parental haplotypes (Pearson's r = 0.2, Fig. 1D), suggesting that differences in effect sizes mostly emerge nearby or within imprinted regions. 664 hq-CpGs (94.5%) out of 702 that exhibited parent-of-origin specific age association and significant difference in effect sizes, were within 100 kb of previously reported imprinted regions, with a mean distance of 2161 bp[12,19–21] (Supplementary Fig. 5).

Of the 702 hq-CpGs, 78.1% showed age association on the paternal haplotype and 21.9% on the maternal haplotype (Fig. 1A). Notably, 535 hq-CpGs (76.2%) associated with the unmethylated or non-imprinted allele, which is consistent with a commonly observed pattern in imprinted regions where paternal alleles are typically unmethylated and maternal alleles methylated[22]. We typically see a decrease in the methylation difference between the parental haplotypes (Fig. 2A), although instances of increased methylation differences are also observed (Supplementary Fig. 6). 0.075% of the non-imprinted CpG islands (95% CI: 0.052–0.105%) were significantly associated with age ($p < 1.1 \cdot 10^{-6} = 0.05/46{,}931$ CpG islands), while 61.7% of the imprinted CpG islands (95% CI: 59.3–64.1%) is associated with age. Together, these findings support a progressive erosion of parent-of-origin-specific methylation patterns with age in imprinted regions.

Among all imprinted regions, a region found within exon 2 of *DIRAS3* (*DIRAS3:Ex2-DMR*)[12], exhibited one of the largest differences in effect sizes between maternal and paternal haplotype associations with age (Fig. 2A, B). Most imprinted genomic regions contain only a limited subset of hq-CpGs with parent-of-origin specific age associations and are therefore distinct from the broad methylation abnormalities typically observed in imprinting disorders, which most often affect most CpGs across the region. In contrast, *DIRAS3:Ex2-DMR* was unusual in that nearly all CpGs exhibit parent-of-origin specific age-associated methylation changes (Supplementary Fig. 7). *DIRAS3* is a maternally imprinted tumor suppressor gene, which is normally expressed exclusively from the paternal allele[23]. As individuals age, we observed progressive methylation on the paternal allele (Fig. 2C, left panel), while the maternal allele remained methylated (Fig. 2C, right

https://doi.org/10.1038/s41467-026-76240-w



panel). Notably, these changes occur during adolescent development and early adulthood.

Discussion

Downregulation of *DIRAS3* has been reported in cancers, including ovary, breast, brain, prostate and thyroid[23–25]. At the *DIRAS3* locus, we observed progressive methylation gain on the active paternal allele with age, resulting in attenuation of the methylation differences between the parental alleles. However, because *DIRAS3* is not detectably expressed in blood, we cannot directly assess whether these methylation changes are associated with altered gene expression. Previous studies have reported age-associated instability at selected imprinted loci, particularly in brain tissue[10]. Our findings extend these observations by providing a genome-wide assessment of parent-of-origin specific methylation changes across

Fig. 1 | Age-associated DNA methylation changes across the genome with age. **A** Overview of genome-wide age association methylation patterns. Among 17.96 million high-quality CpG units (hq-CpGs), 20.8% associate with age (dark green), of which 7.6% (purple) and 92.4% (pink) showed hyper- and hypomethylation, respectively. 79.2% of the hq-CpGs did not associate with age (orange). 1,274,315 parent-of-origin specific associations, with 45.8% of associations reaching significance thresholds on both alleles (light green) and 54.2% on one allele (yellow). 702 show parent-of-origin specific associations with imprinted methylation, thereof 78.1% associate with the paternal allele (brown) and 21.9% with the maternal allele (gray). Note that the total size of the bars is different, and the size of each bar is written on top. **B** Boxplot showing the relationship between chronological age ($n = 7,284$), binned by 5 years (x-axis) and mean methylation levels (y-axis). The two-sided Pearson correlation coefficient (Pearson's r) is shown in red. **C** Enrichment and depletion of age-associated hq-CpGs ($n = 3,739,132$) across genomic regulatory regions. Orange represents genomic context regions, green regulatory regions, and purple repressive genomic regions. Odds ratios and 95% confidence intervals (CI) are shown for each region class. The dashed red line indicates an odds ratio = 1 (no enrichment). Shown for hypermethylated age-associated hq-CpGs (upper panel) and hypomethylated age-associated hq-CpGs (lower panel), separately. **D** Boxplot showing the relationship between differences in mean methylation levels between parental haplotypes (x-axis) and the absolute difference in effect size (β) between haplotypes (y-axis) for hq-CpGs ($n = 1,274,315$) with parent-of-origin specific age-association. The two-sided Pearson correlation coefficient (Pearson's r) is shown in red. For all boxplots in this panel, the center line indicates the median, the box spans the interquartile range (IQR), and whiskers extend to the most extreme observations within 1.5·IQR of the box.

imprinted genomic regions using long-read sequencing. More broadly, these results illustrate how long-read sequencing can characterize age-dependent methylation changes that would not be detectable using unphased methylation measurements.

Several limitations should be considered when interpreting these findings. First, genomic imprinting is formally defined as parent-of-origin-specific gene expression, whereas our analysis focused on parent-of-origin-specific DNA methylation. Although allele-specific methylation is a key regulatory feature of many imprinted loci, we did not assess allele-specific expression and therefore cannot determine whether the observed methylation changes result in loss of imprinting at the transcriptional level. Second, all analyses were performed in whole blood, and age-associated methylation changes and biological consequences may differ across tissues. In particular, because *DIRAS3* is not detectably expressed in whole blood, we could not assess the functional consequences of the observed methylation changes. Third, although our methylation clock achieved lower prediction error than established methylation clocks in the cohorts analysed here, direct comparison between clocks is challenging because they were developed using different technologies, CpG sets, and study populations. Additional validation across independent cohorts and tissues will therefore be required to establish the generalizability of our clock. Despite these limitations, our study highlights how LRS can uncover the full landscape of age-dependent methylation changes, from genome-wide patterns to parent-of-origin-specific age effects. Future studies integrating LRS across diverse tissues with gene expression data will be needed to determine how these parent-of-origin-specific methylation alterations affect gene expression and cellular function.

By interrogating ~18 million hq-CpGs across the genome, we identified widespread age-related methylation changes, developed an accurate methylation aging clock using LRS data, and demonstrated the ability of LRS to resolve parent-of-origin specific methylation dynamics at scale. Phasing the methylation data to parental haplotypes revealed that age-dependent changes are not only widespread but can also be asymmetric across parental haplotypes. Together, these findings establish nanopore sequencing as a powerful framework for studying age-associated methylation dynamics. The ability to detect both genome-wide and parent-of-origin specific methylation changes associated with age provides new opportunities to understand how epigenetic changes accumulate across the lifespan, to develop improved biomarkers of biological aging, and to investigate how they may influence disease risk.

Methods

Dataset and ethics declarations

The study was approved by the National Bioethics Committee in Iceland (Approval no. VSN 14–015) and conducted in agreement with instructions issued by the Data Protection Authority in Iceland (PV_2017060950ÞS/–). All participating subjects signed informed consent. The personal identities of the participants and biological samples were encrypted by a third-party system approved and monitored by the Data Protection Authority. The study complies with all relevant regulations regarding the use of data from human participants and was conducted in accordance with the criteria set by the Declaration of Helsinki. No statistical method was used to predetermine the sample size.

We nanopore-sequenced DNA isolated from whole-blood samples of Icelanders participating in various studies at deCODE Genetics. 7284 samples (3497 male, 3787 female) passed all quality filters and were used for further analysis. Participants were born between 1876 and 2014, with median birth year of 1960 for males and 1958 for females. 778 of the individuals (369 males, 409 females) were deceased at the time of the study. All individuals gave informed consent, and all personal identifiers were encrypted by an external agent before being imported into deCODE's database. For validation, we used 545 samples from an independent cohort from Intermountain USA (223 males, 322 females) born between 1933 and 2002. The participants were recruited by HerediGene: Population study, a large-scale collaboration between Intermountain Healthcare, deCODE genetics and Amgen, Inc[26]. Biological sex was considered in the study design and statistical analyses where appropriate. Sex was assigned based on the detected sex chromosome complement (XX or XY) derived from the sequencing data. Gender identity was not assessed.

We calculate the chronological age of individuals as years from the date of birth to the date of blood draw (sampling date). Only 31.4% of samples had a sampling date registered. For the remaining samples, we calculated the age as years from the date of registration (when the sample was added to the database), which was usually within a few days of the sampling date (median = 22 days, mean = 310 days). We found 5 individuals with age over 110 years, which were all registered in March or April 2016 (Supplementary Fig. 8). To ensure high-quality training data, we exclude all 53 samples registered in these months from our training set. Additionally, we removed 37 samples with a year of death before the sampling date. These samples were included in the test set. The age ranged from 1-106 years old, with mean age of 50 years for females and 47 for males (Supplementary Fig. 9A).

Sample processing and sequencing

DNA from whole blood was extracted using the Chemagic method (PerkinElmer), an automated procedure that involves the use of M-PVA magnetic beads. Sequencing libraries were generated using the SQK-LSK109 ligation kit from ONT. Sample input varied from 1 to 5 μg DNA, depending on the exact version of the preparation kit and the flowcell type used for the PromethION sequencing. Samples were loaded onto PromethION R9.4.1 flowcells following ONT standard operating procedures. Sequencing was performed on PromethION devices. Basecalling was performed using guppyrh-5.0.11 and reads were aligned to reference genome GRCh38[27] using minimap 2 (v. 2.22-r1105-dirty)[28]. For quality control, flowcells with error rate greater than or equal to 15% ($n = 37$), with sequencing coverage less than or equal to 3x ($n = 1$), n50 less than or equal to 7000 bp ($n = 16$) and mean methylation levels lower than or equal to 0.5 ($n = 2$) were removed. Mean sequencing

https://doi.org/10.1038/s41467-026-76240-w

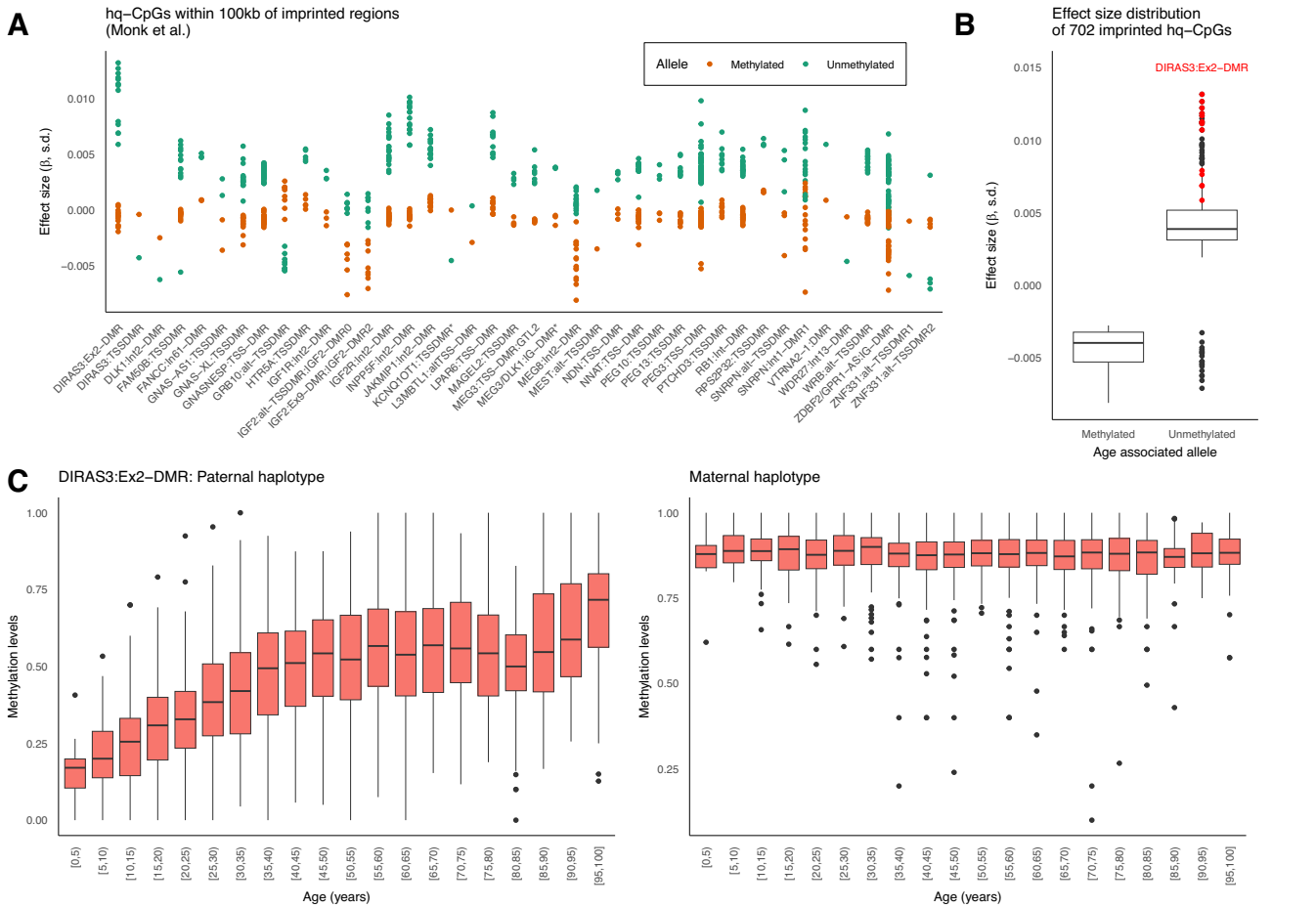


Fig. 2 | Parent-of-origin specific methylation associated with age within imprinted regions. A Distribution of standardized effect sizes (β, s.d. y-axis) for the 702 high-quality CpG units (hq-CpGs) that showed significant parent-of-origin specific age association and were located within 100 kb of 40 imprinted regions listed out in Monk et al.[12] (x-axis). Methylated hq-CpGs are shown in orange and unmethylated in green. Most imprinted regions contain only a subset of hq-CpGs with parent-of-origin specific age associations, whereas the *DIRAS3:Ex2:DMR* was unusual in that nearly all hq-CpGs within that region exhibit this pattern. Y-axis is expressed in standard deviation (s.d.) units. **B** Box plot showing standardized effect sizes (y-axis) for methylated and unmethylated alleles ($n = 702$). Hq-CpGs within the imprinted region at *DIRAS3:Ex2:DMR* are shown in red. Y-axis is expressed in s.d. units. The center line indicates the median, the box spans the interquartile range (IQR), and whiskers extend to the most extreme observations within 1.5·IQR of the box. **C** Boxplot showing methylation levels for hq-CpGs with 5x coverage or more (y-axis) of the paternal (left panel) and maternal haplotype (right panel) across age, binned by 5 years (x-axis) in *DIRAS3:Ex2:DMR* for 7,284 individuals. For all boxplots in this panel, the center line indicates the median, the box spans the interquartile range (IQR), and whiskers extend to the most extreme observations within 1.5·IQR of the box.

coverage was 15.9x and mean n50 was 18,951 bp (Supplementary Fig. 9B, C). This resulted in 9475 flowcells used for 7829 samples (from Iceland and Intermountain).

DNA methylation detection with Nanopolish

All flowcells were methylation called using Nanopolish[29], version 0.13.3. CpGs were measured as units if they were located within 10 base pairs (bp) of each other, and consequently, we refer to them as CpG units. All CpGs within CpG unit therefore are assigned the same methylation status. In short, Nanopolish assigns a log-likelihood ratio for each CpG unit in a read. We interpret values above 1.921 as 5mC methylation and lower than −1.921 as unmodified C. We exclude ambiguous methylation predictions ($-1.921 < LLR < 1.921$).

We define a set of hq-CpGs[14], by calculating strand difference and fraction of non-ambiguous reads out of all reads and removing CpGs with strand difference less than 0.2 and fraction of non-ambiguous reads out of all reads over 0.5. Different from our previous paper[14], we include CpGs within dark regions and CpGs within 5 base pairs of common variants. Instead, we set the methylation levels within 5 bp of common variant, as missing values for the carriers (Supplementary Fig. 10). 2,926,670 hq-CpGs were within 5 bp of 3,586,779 variants, with median number of carriers of 21 (Q1 = 6, Q3 = 114).

Finally, we compute methylation levels as $n_{methylated}/(n_{unmethylated} + n_{methylated})$, where $n_{unmethylated}$ and $n_{unmethylated}$ are the number of reads deemed methylated and unmethylated, respectively.

Principal components

To account for potential batch effects and other latent variables, we computed principal components (PCs) based on a random subset of 200,000 hq-CpGs. PCs were derived separately for each autosome using a leave-one-chromosome-out approach to avoid circularity. PCs were computed using age and sex adjusted methylation data, for each flowcell using the PCA function from sklearn.decomposition in Python (v.3.11.7). To calculate variance explained, we calculate PCs over all autosomes. The top 10 PCs, selected based on the proportion of variance explained (Supplementary Fig. 1), were included as covariates in all regression models.

These components capture major sources of variation, including blood cell composition, sampling quality, and sequencing batch effects[15]. Finally, for each sample, we derived a sample-level PC profile by averaging across all associated flowcells, which yields similar results as per flowcell (Supplementary Fig. 1). These averaged PCs were then used as covariates.

Age association

To assess the relationship between DNA methylation and chronological age, we applied a linear regression model, implemented with lm in R (v.4.6.1). For each CpG unit, we rank first normal transformed methylation levels ($CpG_{rank\,normal}$) and modeled as a function of age, with additional covariate adjustment for sex, the first ten PCs to account for technical and biological variability and genotypes of strongest cis-acting ASM-QTL ($genotype_{best\,marker}$).

The model was specified as:

$$CpG_{rank\_normal} \sim age + sex + PC_1 + \cdots + PC_{10} + genotype_{best\_marker} \quad (1)$$

We assessed statistical significance using p-values from linear regression and reported effect sizes (β) as regression coefficients. Positive effect size indicates hypermethylation, while a negative effect size indicates hypomethylation. To control for multiple testing, we applied Bonferroni correction by adjusting the significance threshold based on the total number of tests performed. To validate that our results are not an artifact of our data transformations, we also calculate association without covariates and without standardizing methylation values, and compared the p-values and direction of effect sizes.

We calculate residual variance in methylation (or root mean squared error, RMSE) for each hq-CpG unit with the following formula.

$$RMSE = \sqrt{\frac{\sum (y_i - \tilde{y}_i)^2}{n}} \quad (2)$$

Where $(y_i - \tilde{y}_i)^2$ are the squared residuals and n is total number of samples.

No CpG-level filtering based on coverage was applied, apart from the filter applied when selecting hq-CpGs[14].

1934 individuals had blood cell counts available from the same blood draws as were used to isolate the DNA for nanopore sequencing. We included counts of lymphocytes, basophils, eosinophils, immature granulocytes, monocytes, white blood cells, and red blood cells, and the fraction of nucleated red blood cells. We did not include neurophils counts because of their strong correlation with lymophocyte counts (Pearson's $r = -0.94$) to avoid colinearity in the regression.

Similarly, we perform parent-of-origin specific age associations by looking at each haplotype as separate measurement and specify the model as:

$$CpG_{rank\_normal} \sim age + sex + haplotype + PC_1 + \cdots + PC_{10} + genotype_{best\_marker} \quad (3)$$

Where haplotype was either paternal or maternal.

Evaluating robustness of hq-CpGs age associations

To evaluate whether the observed associations between DNA methylation at CpG units and chronological age could have arisen by chance, we performed a permutation test. Specifically, we tested the null hypothesis that methylation is not associated with age by randomly permuting the age variable across individuals and recalculating the association statistics for each hq-CpG unit. For each of the 17,959,684 hq-CpGs, we repeated the age association analysis ten times with the age values randomly permuted across individuals in each iteration, thereby breaking any real association while preserving the distribution of both variables. In each permutation, we recalculate the p-values from the linear models and counted the number of hq-CpGs with significant association, yielding an empirical null distribution of the number of significant hq-CpGs under the assumptions of no association.

Genomic enrichment analysis

All genomic regions were defined from the GRCh38 reference genome[27]. We used annotations from encode ChromHMM (18-state model)[30] and selected BSS01380_NEUTROPHIL cell line for all analysis. For validation we did the same analysis for the BSS00178_CD14_MONOCYTE cell line (Supplementary Fig. 11), yielding similar results. For genomic context (coding exons, exons, intergenic regions and introns) we used GENCODE (v48)[31]. We split the genome into 100 kb windows. From each window we select one CpG and check whether the CpG is age-associated and whether it was hyper- (effect size, $\beta \geq 0$) or hypomethylated (effect size, $\beta < 0$). We then check whether the CpG is within the genomic region and create two contingency tables, for hyper- and hypomethylated counts separately. We used Fisher's exact test to calculate odds-ratios and p-values from the contingency tables.

Methylation phasing

As described in our previous study[15], DNA sequence variants were called using Graphtyper[32] based on 63,460 Illumina whole-genome sequenced individuals representing a subset of 173,025 SNP chip-typed individuals from the Icelandic population. The sequenced individuals were then phased to impute haplotypes and assign parent of origin into long-range phased chip-typed individuals[33].

The long-range phasing and assigned parent of origin, enabled us to assign nanopore sequences to maternal or paternal chromosomes[15]. Parental haplotypes were assigned by examining the phasing status of a set of 8,960,728 high-quality sequence variants, using heterozygous carriers of in-read sequence variants, allowing us to assign CpG methylation calls (methylated or unmethylated) to maternal and paternal chromosomes. Reads overlapping at least three heterozygous variants and where at least 70% of the variants were consistent with the phasing of one parent, were considered phased and used for subsequent analysis.

For each hq-CpG unit and each parental haplotype, we defined the haplotype methylation levels as the number of sequences of that parental haplotype that are methylated at the hq-CpG unit divided by the total number of parental sequences with reliable methylation calls covering the hq-CpG unit.

Difference in effect sizes

We evaluate whether there is significant difference between the effect sizes (β) of the parental haplotypes using z-test by first calculating z-score using the formula:

$$z = \frac{\beta_M - \beta_P}{\sqrt{SE_M^2 + SE_P^2}} \quad (4)$$

Where β_M is the effect size of the maternal haplotype and SE is the standard error. We then find the corresponding p-value and perform a Bonferroni correction.

Defining imprinted CpG islands and the midpoint of imprinting

We looked at the methylation-depleted sequences (MDS[15]) and found which MDS overlaps CpG-islands from hg38 annotation[27]. We then checked whether the CpG islands were within MDSs that we found to be overlapping any of the imprinted regions from Monk et al., Court et al. and Joshi et al. We calculated the middle of an imprinted region as the sum of start position and half of the length of the region.

Construction and validation of a methylation aging clock

We use all 7284 samples, from the Icelandic cohort, for either testing or training, split into 80% training (5,828 samples) and 20% testing (1625 samples). We ensure high-quality training set by excluding samples with registration dates in March or April 2016 and with year of

birth after sampling date, as this was an indication of error in registration. We further limit our training set to individuals between 18-100 years old. Combined, these criteria excluded 477 samples from the training set, but these samples were used for testing; the remaining 1148 samples were randomly selected into the testing set.

We used createDataPartition() in R, stratified by age group and sex, to ensure balanced representation, to split the remaining 6807 samples randomly into test and training. The training set consisted of 3030 females and 2798 males, while the test set consisted of 680 females and 945 males.

We first perform feature selection of hq-CpGs using linear regression between methylation levels in CpGs and chronological age and select 1 million hq-CpGs with lowest p-values. We excluded hq-CpGs units with more than 80% missing values and for the remaining hq-CpGs we mean imputed missing values, using SimpleImputer(-strategy = "mean") from the scikit-learn library in python. Subsequentially, all features were standardize using StandardScaler, also from scikit-learn.

We implemented a Lasso regression model with 5-fold cross-validation, using GridSearchCV() from the Python package scikit-learn, to tune the regularization parameter. We evaluated the performance using the negative mean squared error metric. We trained the final model on the full training dataset using the optimal alpha and included features (hq-CpGs) selected by the Lasso model, and evaluated the model performance on both the remaining samples (test) and an independent validation cohort from Intermountain. Metrics reported include root mean squared error (RMSE), mean absolute error (MAE), and median absolute error (MedAE).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequence data and the Icelandic genealogical database cannot be made publicly available because Icelandic law and the regulations of the Icelandic Data Protection Authority prohibit the release of individual-level and personally identifying data. Data access for raw data can be granted for scientific purposes only at the facilities of deCODE genetics in Iceland, subject to Icelandic law regarding data usage. Anyone wishing to gain access to the data should contact Bjarni Halldórsson (bjarni.halldorsson@decode.is) or Daníel Guðbjartsson (dfg@decode.is). Requests for access are generally considered monthly. The CpG unit summary statistics and Supplementary Data 1-5 are available from Zenodo: https://doi.org/10.5281/zenodo.21374514[34]. Data from the following publicly available databases were used and are available from:reference genome (GRCh38/hg38): https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/. Genome annotations (Gencode, v.48): https://www.gencodegenes.org/human/ Encode ChromHMM 18-state model: https://www.encodeproject.org/annotations/ENCSR334AGV/. Information on all data generated in this study are provided in the Supplementary Information file.

Code availability

The code used to develop the model//perform the analyses and generate results inthis study was developed using R (v.4.6.1), https://www.r-project.org/ and Python (v.3.11.7 with numpy v.2.3.4, pandas v.2.2.2, scipy v.1.7.2), https://www.python.org/downloads/. This study used used the following publicly available software: GraphTyper (v.2.7.1), https://github.com/DecodeGenetics/graphtyper.

References

1. Ehrlich, M. et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Res* **10**, 10.1093/nar/10.8.2709 (1982).
2. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 10.1038/366362a0 (1993).
3. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 10.1101/gad.2037511 (2011).
4. Wang, K. et al. Epigenetic regulation of aging: implications for interventions of aging and diseases. *Signal Transduct. Target. Ther.* **7**, 374 (2022).
5. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115 (2013).
6. Hannum, G. et al. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol Cell* **49**, 10.1016/j.molcel.2012.10.016 (2013).
7. Levine, M. E. et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
8. Lu, A. T. et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 10.18632/aging.101684 (2019).
9. Bell, C. G. et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol.* **20**, 249 (2019).
10. Mancino, S. et al. Exploring the Stability of Genomic Imprinting and X-Chromosome Inactivation in the Aged Brain. *Aging Biology* **2**, 20240030 (2024).
11. Li, Y. & Sasaki, H. Genomic imprinting in mammals: its life cycle, molecular mechanisms and reprogramming. *Cell Res.* **21**, 466–473 (2011).
12. Monk, D. et al. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat. Rev. Genet.* **20**, 235–248 (2019).
13. Woodfine, K., Huddleston, J. E. & Murrell, A. Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. *Epigenetics Chromatin* **4**, 10.1186/1756-8935-4-1 (2011).
14. Sigurpalsdottir, B. D. et al. A comparison of methods for detecting DNA methylation from long-read sequencing of human genomes. *Genome Biol* **25**, 10.1186/s13059-024-03207-9 (2024).
15. Stefansson, O. A. et al. The correlation between CpG methylation and gene expression is driven by sequence variants. *Nat Genet* **56**, 1624–1631 (2024).
16. Lu, A. T. et al. Universal DNA methylation age across mammalian tissues. *Nat Aging* **3**, (2023).
17. Field, A. E. et al. DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. *Molecular Cell* vol. 71, https://doi.org/10.1016/j.molcel.2018.08.008 (2018).
18. Maegawa, S. et al. Widespread and tissue-specific age-related DNA methylation changes in mice. *Genome Res* **20**, 10.1101/gr.096826.109 (2010).
19. Joshi, R. S. S. et al. DNA Methylation Profiling of Uniparental Disomy Subjects Provides a Map of Parental Epigenetic Bias in the Human Genome. *Am. J. Hum. Genet.* **99**, 10.1016/j.ajhg.2016.06.032 (2016).
20. Court, F. et al. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome Res* **24**, (2014).
21. Zink, F. et al. Insights into imprinting from parent-of-origin phased methylomes and transcriptomes. *Nat Genet* **50**, (2018).
22. Moore, G. E. et al. The role and interaction of imprinted genes in human fetal growth. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 370, https://doi.org/10.1098/rstb.2014.0074 (2015).
23. Bildik, G. et al. DIRAS3: An Imprinted Tumor Suppressor Gene that Regulates RAS and PI3K-driven Cancer Growth, Motility, Autophagy, and Tumor Dormancy. *Molecular Cancer Therapeutics* vol. 21, https://doi.org/10.1158/1535-7163.MCT-21-0331 (2022).
24. Yu, Y. et al. Biochemistry and Biology of ARHI (DIRAS3), an Imprinted Tumor Suppressor Gene Whose Expression Is Lost in Ovarian

and Breast Cancers. *Methods in Enzymology* vol. 407, https://doi.org/10.1016/S0076-6879(05)07037-0 (2006).
25. Soboska, K. et al. Expression of RASSF1A, DIRAS3, and AKAP9 Genes in Thyroid Lesions: Implications for Differential Diagnosis and Prognosis of Thyroid Carcinomas. *Int. J. Mol. Sci.* **25**, (2024).
26. Taylor, D. P. et al. HerediGene Population Study IT infrastructure: A model to support genomic research recruitment and precision public health. *AMIA Annu. Symp. Proc.* 2023, (2023).
27. Schneider, V. A. et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 10.1101/gr.213611.116 (2017).
28. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 10.1093/bioinformatics/bty191 (2018).
29. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* **14**, 407–410 (2017).
30. Boix, C. A., James, B. T., Park, Y. P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 10.1038/s41586-020-03145-z (2021).
31. Frankish, A. et al. GENCODE 2021. *Nucleic Acids Res* **49**, 10.1093/nar/gkaa1087 (2021).
32. Eggertsson, H. P. et al. Graphtyper enables population-scale genotyping using pangenome graphs. *Nat Genet* **49**, 10.1038/ng.3964 (2017).
33. Kong, A. et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet* **40**, 1068–1075 (2008).
34. Sigurpalsdottir, B. Nanopore sequencing identifies parent-of-origin specific age-associated methylation changes at imprinted loci in the human genome. Zenodo. https://doi.org/10.5281/zenodo.21374514 (2026).

Acknowledgements
We thank colleagues at Amgen deCODE genetics for helpful discussions and feedback.

Author contributions
B.D.S., O.A.S., and B.V.H. designed the study. B.D.S and B.V.H. implemented the software. B.D.S., O.A.S., and B.V.H. analyzed the data and interpreted the results. G.H. and S.Þ.S. implemented the sequencing pipeline. P.I.O. and A.G. performed the long-range phasing of the variants, supervised by G.M. O.Þ.M. and D.N.S. performed the nanopore sequencing. B.D.S. wrote the initial version of the manuscript, and B.V.H. and O.A.S. contributed to the subsequent versions. All authors contributed to the final version of the manuscript.

Funding
The authors received no specific funding for this work.

Competing interests
All authors are employees of Amgen deCODE genetics. B.D.S, B.V.H, G.H., and O.Þ.M. have received travel funds to speak at events hosted by Oxford Nanopore.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-76240-w.

Correspondence and requests for materials should be addressed to Brynja Sigurpalsdottir or Bjarni V. Halldorsson.

Peer review information *Nature Communications* thanks Robert Feil and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.