Genome Biology

RESEARCH

Open Access

Sex-specific nonlinear DNA methylation aging trajectories reveal biomarkers of cancer risk and inflammation

Robin Grolaux[1], Macsue Jacques[1], Bernadette Jones-Freeman[1], Steve Horvath[2], Andrew Teschendorff[3] and Nir Eynon[1*]

*Correspondence:
Nir.Eynon@monash.edu

[1] Australian Regenerative Medicine Institute, Monash University, Clayton, VIC 3800, Australia
[2] Altos Labs, Cambridge, UK
[3] CAS Key Lab of Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai, China

Abstract

Background: Aging is a multi-modal process, leaving distinct molecular signatures across the epigenome. DNA methylation is among the most robust biomarkers of biological aging, yet most studies assume linear age relationships and analyze mixed-sex cohorts, overlooking known sex differences. Such approaches risk obscuring critical nonlinear transitions and sex-specific trajectories.

Results: We develop SNITCH, a computational framework to detect complex nonlinear methylation trajectories and disentangle shared from sex-divergent patterns. Applied to the array-derived whole-blood methylomes from 252 females and 246 males (ages 19–90 years), SNITCH reveals convergent and divergent epigenetic aging pathways independent of immune cell composition. Nonlinear trajectories are enriched for developmental transcription factor motifs, including NF1/CTF and REST, with known oncogenic roles. Importantly, a female-specific nonlinear cluster is prospectively associated with cancer onset and systemic inflammation in an independent cohort, nominating clinically relevant biomarkers. We replicate the analysis in an additional cohort and highlight consistent nonlinear trajectories.

Conclusions: Our results uncover sex-specific, nonlinear aging programs that capture the dynamics of epigenetic change beyond linear models. These findings provide potential candidate biomarkers for early disease risk and advance understanding of how aging trajectories diverge between sexes.

Keywords: Aging, Nonlinear, DNA methylation, Sex differences, Epigenetic, Biomarkers, Computational biology

Background

Biological aging is often modeled as a linear process; yet many molecular and physiological changes accelerate, decelerate, or shift phases with age rather than progressing uniformly [1]. This concept of biological nonlinearity, changes that deviate from a constant rate over time, is increasingly supported across diverse molecular and

physiological domains. Telomere attrition, a hallmark of aging, follows a nonlinear trajectory, with faster shortening in early life and slower decline in later decades [2]. Transcriptomics studies in mice have identified cross-tissue nonlinear gene expression patterns that correlate with protein expression [3], as well as late-life shifts in skeletal muscle expression consistent with the 'elbow' (i.e. inflection point), typical of nonlinear functions [4]. In humans, age-related nonlinear patterns have been reported for circulating microRNAs [5], and undulating changes have been observed in the proteome [6], transcriptome [7], and metabolome [1]. Together, these findings suggest that many molecular processes follow complex trajectories across the lifespan, challenging the notion of steady, uniform aging.

DNA methylation (DNAm) is a primary hallmark of aging, and one of the most reliable molecular surrogates for estimating biological age [8, 9]. While many epigenetic clocks appear to have a linear relationship with chronological age, their underlying regression models often assume a more complex, log-linear relationship: notably, the Horvath 2013 pan-tissue clock and the Lu 2023 pan-mammalian clock 3 were built by regressing DNA methylation data on a log-transformed version of age for samples < 20 years old, while an untransformed linear model was used for older samples, which suggests that methylation changes rapidly in early life and then slows to a constant rate after adulthood [10–12]. While these regression modeling approaches recognize the nonlinear nature of methylation aging, they constrain trajectories to simple monotonic forms and cannot capture more complex patterns such as U-shaped curves, multi-phase dynamics, or abrupt inflection points. Indeed, lifespan analyses have documented nonlinear DNAm changes at specific CpG sites [13–16]. Measurements of increased variance, such as variably methylated positions, also exhibit nonlinear age-related patterns, highlighting that both the mean and variability of DNAm can shift in complex ways over time [14, 17, 18]. Despite this, most DNAm studies continue to rely on linear regression or other monotonic models, which preferentially detect features that increase or decrease steadily, while potentially missing, or mischaracterizing, nonlinear signals [19].

Sex-specific differences in aging are widely recognized at physiological and clinical levels [20–22], yet remain underexplored in the context of DNAm. When considered, sex is most often modeled as an interaction term in linear frameworks, limiting the detection of patterns that differ in shape, timing, or magnitude between sexes [1, 15, 21, 23]. Nonlinear approaches are particularly well-suited to uncover such differences, as they can reveal age windows of abrupt divergence, sex-specific inflection points, or distinct multi-phase dynamics.

Here, we investigated sex-specific nonlinear aging patterns of DNAm across the adult human lifespan. To enable this, we developed SNITCH (Semi-supervised Nonlinear Identification and Trajectory Clustering for High-dimensional data), a robust framework for detecting and clustering CpG sites with shared linear and nonlinear age-related changes. We first validated SNITCH using simulated data, then applied it to a whole-blood DNAm dataset (EPIC array; $N = 238$ males, and 256 females; 18–90 years old), accounting for immune cell composition. This analysis identified both sex-dependent and sex-independent nonlinear aging trajectories. Replication in an independent cohort highlighted conserved nonlinear CpGs and aging trajectories, with specific nonlinear clusters associated with inflammation and cancer onset in a sex-specific manner.

Results

SNITCH: semi-supervised approach to cluster CpGs based on their aging pattern

Given the widespread evidence of nonlinear aging dynamics and the shortcomings of linear analyses, there is a clear need for new methods to detect and characterize nonlinear patterns of aging. Understanding aging's complex trajectory requires analytical approaches that can capture inflection points, accelerations/decelerations, and multiphase changes in biological data. By moving beyond linear models, researchers can unveil previously hidden aging signals. Previous attempts to identify nonlinear changes in DNAm have relied on the binning of age categories [15] or a priori shape of aging patterns [14]. However, systematic tools to scan for arbitrary nonlinear trajectories, without pre-specifying a particular model, are needed to truly let the data reveal aging's patterns. The closest attempt to answer this need has been described by Okada et al. [19] where functional data analysis was used to cluster CpGs as linear increasing (LI), linear decreasing (LD), non-correlated (NC), or nonlinear (NL). Nevertheless, this method remained limited in its ability to discriminate between nonlinear trajectories and identify increased variance (VI).

To address this gap, we developed and applied SNITCH (Semi-supervised Nonlinear Identification and Trajectory Clustering for High-dimensional data), a heuristic-based statistical framework that distinguishes between linear, nonlinear, variable, and non-correlated methylation trajectories. The method leverages both generalized linear modeling and generalized additive modeling to identify CpGs exhibiting distinct age-associated patterns (NC, LI, LD, NL, and VI) while controlling for potential confounders. Unsupervised clustering is then applied on functional principal components of the nonlinear positions to highlight CpGs sharing similar nonlinear trajectories (Fig. 1A). We benchmarked the classification accuracy of SNITCH, in combination with clustering of nonlinear patterns, against three stand-alone unsupervised clustering algorithms and the functional trajectory-based DICNAP method [19]. Performance was evaluated using Adjusted Rand Index (ARI) and Adjusted Mutual Information (AMI), using simulated methylation data (i.e., bound by 0–1) from a highly varied pool of distributions with known ground-truth labels (Fig. 1B, Methods). We found that SNITCH outperformed stand-alone unsupervised clustering algorithms and DICNAP (Fig. 1D, Additional file 1: Fig. S1A). The best results were achieved by using SNITCH + Fuzzy/HDBSCAN (ARI: 0.97; AMI: 0.98). In this setting, we observed a robust concordance between predicted and ground truth labels, with the main misclassifications occurring between the logarithmic decreasing and linear decreasing groups (Fig. 1C, E).

Sex-specific nonlinear aging patterns in blood

We applied SNITCH to a blood dataset (GSE246337) containing DNAm data generated on the EPICv2 array for 238 and 256 males and females (age: 18–90 yo) (Fig. 2A). Blood DNAm is highly influenced by immune cell heterogeneity, and adjustment for cell type composition is essential for the identification of epigenetic modifications independent of the immune profile [24]. Immune cell fractions of whole blood can be effectively estimated by the use of DNAm-based deconvolution methods [25]. Thus, in both females and males, we built three different models of aging trajectories accounting for an increasing number of immune cell types. Our baseline model didn't include any cell

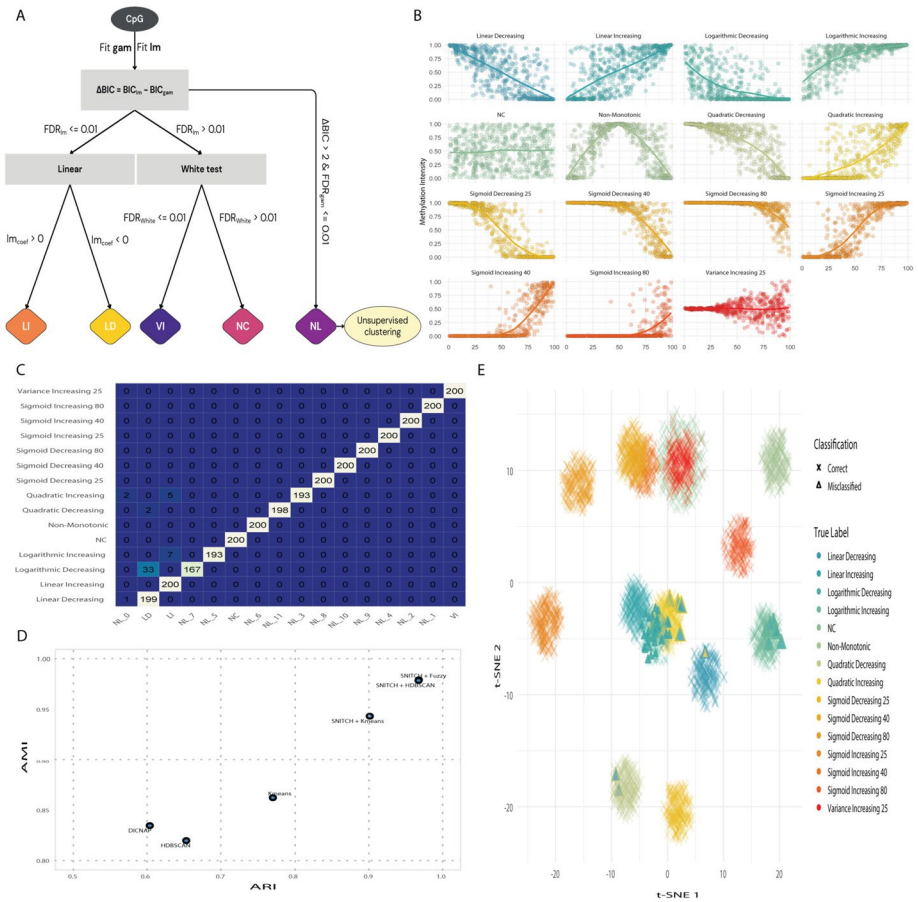

Fig. 1 **A** SNITCH Pipeline. **B** Distribution of the simulated aging patterns. **C** Confusion matrix between the ground truth and predicted clusters. **D** Benchmark of SNITCH compared to stand-alone unsupervised clustering methods and DICNAP. **E** T-SNE representation of the simulated CpGs and their classification. GAM = Generalized Additive Model; lm = Linear Model; BIC = Bayesian Information Criterion; FDR = False Discovery Rate; LI = Linear Increasing; LD = Linear Decreasing; VI = Variance Increasing; NC = Non-Correlated; NL = Nonlinear

fractions. The 7 cell-model was corrected for B-, NK-, CD4T, and CD8T-cells, Monocytes, Neutrophils, and Eosinophils. Finally, the 12-cell model discriminated between naive and mature B-, CD4T-, and CD8T-cells, and added T-regulatory cells and Basophils. This rigorous approach allowed us to identify DNAm aging patterns occurring independently from changes in immune cell fractions.

Focusing on autosomal chromosomes, CpG classification was highly stable across models in both sexes: In females, 95.9% ($N = 534{,}132$) retained identical cluster assignments across all three models. This percentage changed only a little in males (95.6%). This was mainly driven by CpGs classified as Non-Correlated (NC) with age (94% in both males and females). Among the remaining CpGs that changed classification, transitions were most frequently directed toward the NC cluster upon inclusion of immune covariates in both females and males (Fig. 2B, Additional file 2: Table S1), highlighting how linear and nonlinear trajectories are both confounded by changes in the immune profile. The following analyses describe the results from the 12-cell model. In both males and females, the majority of CpGs showed no correlation to age (N_fem = 543,972

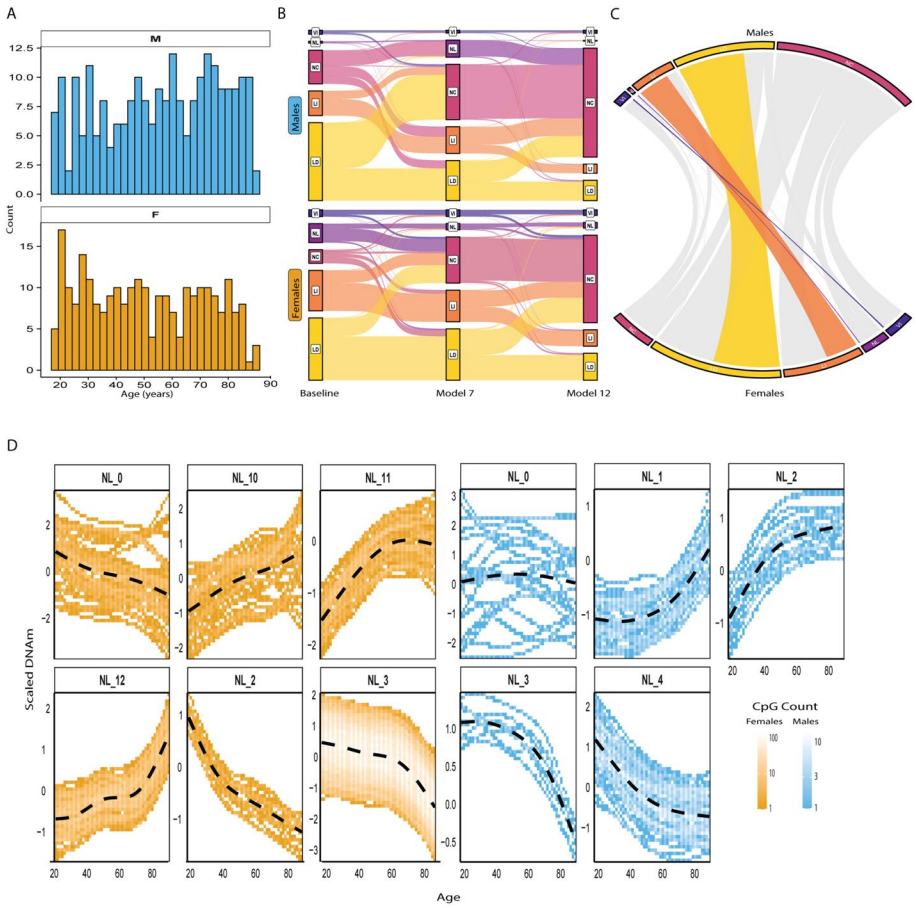

Fig. 2 **A** Age distribution among the cohort. **B** Repartition of the number of CpGs among clusters in females and males. **C** Conserved CpGs between male and female clusters. **D** Nonlinear clusters identified in females and males. Beta values were centered and scaled prior FPCA and unsupervised clustering. LI = Linear Increasing; LD = Linear Decreasing; VI = Variance Increasing; NC = Non-Correlated; NL = Nonlinear. Note: In A & B, CpGs that were NC across all 3 models were removed for visualization purposes

(97.7%); N_m = 548,935 (98.6%)). Whereas the number of Nonlinear (NL) CpGs was markedly different between females ($N = 1305$) and males ($N = 155$) (Fig. 2C, Additional file 1: Fig. S1B). Those results seemingly contrast with a recent meta-analysis showing that almost half of the blood CpGs showed differential methylation with age [18]. We investigated whether our analysis was underpowered by combining male and female participants, effectively doubling the size of the cohort, and accounting for sex as a covariate in the model. Consistent with our sex-specific analysis, we found that 93% of the CpGs ($N = 516{,}836$) stayed non-correlated with age, indicating that our results are stable within the scale of the current cohort. However, the large discrepancy in sample size compared to the meta-analysis remains the most likely explanation for the lower number of age-associated CpGs detected in our study. Our stricter QC (Methods) and accounting for 12 immune cell fractions in the model likely further reduce the number of significant hits.

To evaluate the concordance of CpG aging trajectory classifications between sexes, we compared SNITCH-assigned trajectory labels in males and females using a Chi-square

test of independence. The analysis revealed a highly significant association ($\chi^2 = 403{,}770$, df = 16, $p$value $< 2.2 \times 10^{-16}$), indicating that CpGs were classified into the same trajectory category in both sexes more frequently than expected by chance. Out of the total CpGs assessed, only 9,938 CpGs (~ 1.78%) changed trajectory class between males and females, confirming a high degree of overall consistency. This was further supported by the standardized residuals (Supp. Figure 1C), which showed strong positive values along the diagonal of the contingency matrix, reflecting substantial overlap in classification across sexes, including in the NL–NL cell, indicating an enrichment of the 39 CpGs classified as nonlinear in both sexes despite the discrepancy in the total number of NL CpGs identified. Conversely, the most pronounced negative residuals were observed in off-diagonal cells where CpGs were classified as age-associated (LD or LI) in one sex but NC in the other, suggesting a subset of CpGs with potential sex-specific sensitivity to age-related methylation changes. In addition, NL CpGs showed moderate positive residuals when aligned with LD (+ 140.3) and LI (+ 36.0) in the opposite sex, suggesting that a subset of CpGs classified as nonlinear in one sex may appear more linear in the other. Finally, the NL–NC cells exhibited a residual of −138.4 in females and −61.2 in males, indicating that CpGs classified as nonlinear in one sex were rarely non-correlated in the other, further supporting their functional relevance. To identify clusters of CpGs sharing similar aging trajectories, we applied the last step of our pipeline by performing unsupervised clustering on the functional components of the NL CpGs (Methods). This revealed similar-shaped trajectories in males and females, with 6 primary clusters identified in females and 5 in males (Fig. 2D). To avoid redundancy between similar clusters, we merged clusters with a Spearman correlation coefficient > 0.9 (Additional file 1: Figs. S2A, B and 3A, B). This resulted in a final number of 4 principal NL patterns in females and males (Additional file 1: Figs. S2C, 3C). Within those, we observed different inflection points, marking a change of pace in the methylation trajectory. In females, clusters 3, 11, and 12 showed an elbow between the 70–80 years mark, where cluster 2 showed it earlier, around the 50 years mark. In males, the inflection points appeared around 60 years for clusters 1 and 3 and 50 years for clusters 2 and 4. This analysis highlighted the similar aging patterns seen in males and females, but hinted at different temporalities for the inflection points.

Functional analysis of age-related trajectories

Trajectories of clocks' CpGs

The first step we took to understand the functional role of the clusters we identified was to investigate the classification of CpGs previously used to train epigenetic clocks. A ubiquitous tool in the field of aging, epigenetic clocks are biomarkers that show relevance in assessing the onset of several age-related conditions as well as the utility of therapeutic strategies. Most epigenetic clocks are machine-learning models based on linear regression [26] (e.g., elastic-net). By construction, this limits the granularity of the aging trajectories they capture by either overlooking nonlinear patterns or over-simplifying them as linear, thus limiting their biological interpretability. Further complicating their interpretation, blood-based epigenetic clocks often ignore cell-type heterogeneity when considering age-related DNAm changes. Within our two cohorts, we looked at the classification labels of the CpGs underlying 9 of the most common clocks [11, 27–32]

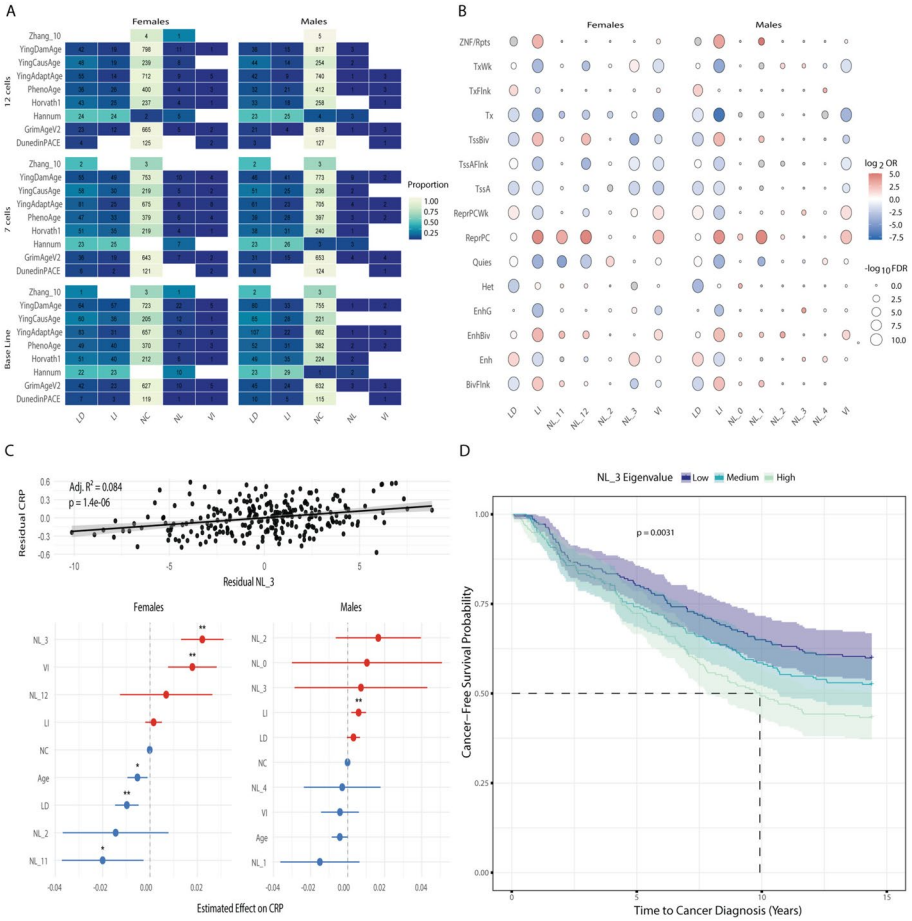

Fig. 3 **A** Enrichment of CpG labels across 9 epigenetic clocks. **B** Chromatin state enrichment analysis across age-related clusters identified in females and males. **C** Top: Regression model of the female NL3 cluster eigenvalue against estimated CRP levels. Bottom: Coefficient from a linear regression model predicting estimated CRP levels with each cluster's eigenvalues in males and females. **D** Kaplan–Meier survival curves stratified by NL3 cluster tertiles in females. Participants with high NL3 scores had a significantly shorter cancer-free survival period compared to those in the low and mid tertiles (log-rank $p = 0.003$). The dashed lines represent the median time before diagnosis. Shaded regions indicate 95% confidence intervals. Abbreviations in B correspond to the classes from the Roadmap Epigenomics project [35]. CRP: C-Reactive Protein

across our 3 models (Fig. 3A). As expected, we found that the proportions of CpGs classified as NC increased between each model iteration, reflecting that part of the signal captured by these clocks arises from age-driven changes in immune-cell fractions [8, 26]. In addition, our results highlighted that the majority of the clocks' CpGs remained classified as NC in our baseline model. Notably, the Hannum clock was the only clock that showed a higher proportion of CpG associated with age (LI & LD) compared to NC across all three models. This finding is consistent with the inherent designs of each clock: the 2nd-generation clocks (PhenoAge, GrimAge, DunedinPace, Zhang_10, and YingAge) were trained on phenotypic age or mortality risk [27, 29–33], overlooking chronological CpGs, and Horvath's clock was trained across multiple tissues [11], where Hannum's clock was built to estimate chronological age in blood [28]. Finally, we observed that most of the clocks captured NL and VI CpGs in males and females. To complement this

analysis, we examined a published set of 350 age-associated CpGs shared across immune cell types [34]. Applying the same approach to these loci yielded classifications of hyper- and hypo-methylation that aligned with those reported in the original study, reinforcing the validity of our modeling strategy, which incorporates immune cell composition (Additional file 1: Fig. S4A). Moreover, our results revealed that a subset of these CpGs exhibited nonlinear associations with age.

Functional analyses of sex-specific aging patterns

After successfully identifying sex-specific aging methylation patterns, we performed separate functional analysis in males and females for all age-associated clusters.

Chromatin enrichment analysis

We performed chromatin state enrichment analysis in males and females to determine the epigenetic context of age-associated methylation clusters. Each cluster was tested for enrichment across 15 chromatin states from the Roadmap Epigenomics Project [35] using the NC cluster as a reference (Methods). Overall enrichment results were highly similar between males and females in the LD, LI, and VI clusters (Fig. 3B, Additional file 3: Table S2). Those similarities are consistent with the overlap of CpGs observed in those clusters between males and females (Fig. 2C). We observed an enrichment of the repressed Polycomb and bivalent chromatin states in clusters characterized by increased methylation or age-related variance (LI, VI, NL11, NL12 in females & LI, VI, NL0, NL1 in males). This enrichment pattern of age-related hypermethylated sites is well known [36–38], and shows that chromatin states at NL clusters are mostly driven by the directionality of the trajectories (gain vs. loss of methylation) rather than by the individual patterns. A noticeable deviation from this was the significant enrichment of "weak transcription" states in NL3 in females (hypomethylated with age) compared to its depletion in LD (NL3 $\log_2$ OR = 0.7, LD $\log_2$ OR = −0.14). This is supported by previous studies highlighting that hypomethylated sites are enriched in active or transcribed genomic regions, including the weak transcription chromatin state [39, 40]. In particular, this suggests a breaking point around 60 years old in females, where erosion of DNA methylation in weakly transcribed or formerly silent chromatin regions could lead to leaky transcription, supporting the theory of increased transcriptional noise in aging [41].

Pathway enrichment analysis

Next, we performed pathway enrichment analysis for our nonlinear clusters using different databases (the Gene Ontology (GO) Molecular Function (MF) and Biological Process (BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG)). Enrichment analyses in DNA methylation datasets are inherently biased toward long genes, and due to CpGs mapping to multiple genes [42, 43]. To account for those, we used *missmethyl,* a method that addresses these biases by leveraging prior probabilities [43, 44]. Across all NL clusters, only NL12 in females was enriched for the term "Neuroactive ligand signaling" in KEGG (FDR = 0.023), a term previously associated with age-related changes in blood [45], no enrichment was observed for NL clusters in males (Additional file 4: Table S3).

Motif enrichment analysis

In addition to the local chromatin context, the underlying DNA sequence also contains biologically relevant information that can help understand the function of a particular set of CpGs. Although the mechanisms haven't been fully elucidated yet, it is widely accepted that the methylation context at a specific motif can positively or negatively impact the binding affinity of transcription factors (TFs) [46, 47]. Focusing on the NL CpGs, we performed cluster-wise motif enrichment analysis to identify the presence of TFs binding sites (TFBs) in their vicinity (Methods). In females, only NL12 and NL3 showed significantly enriched motifs (Additional file 1: Fig. S4B, Additional file 4: Table S3). NL12 was enriched in ZNF652 binding site (qval = 0.0249) where NL3 was enriched for both Nuclear Factor 1 (NF1/CTF) half- (qval = 1×10^{-5}) and full motif (qval = 0.0026), Hoxc9 (qval = 0.0411) and Gata6 (qval = 0.0476). In males, only NL1 and NL4 were enriched in TFBs (Additional file 1: Fig. S4C, Additional file 5: Table S4). Similar to NL3 in females, the NL4 cluster in males was enriched for NF1 half- (qval = 0.0063) and full-site (qval = 1×10^{-4}) along with REST/NRSF (qval = 1×10^{-5}). NF1-CTF factors are widely expressed regulators of gene expression and tissue development, and their dysregulation has been implicated in diverse cancers [48–51]. Similarly, ZNF652 acts as a potent tumor suppressor in breast cancer by repressing the transcription of oncogenes [52–54], while both Gata6 and Hoxc9 are involved in the development and with known oncogenic properties [55–58].

Nonlinear clusters as biomarkers of diseases

Cancer risk

DNA methylation, or surrogate measures such as epigenetic age-acceleration, have been used as biomarkers to predict the onset of or diagnose a wide range of pathological conditions, ranging from rare developmental diseases to cancer [59–64]. Given the presence of transcription factor binding motifs known to regulate oncogenic pathways within the methylation clusters, we hypothesized that these epigenetic modules may capture pre-diagnostic signals of cancer risk. To evaluate the predictive value of the NL clusters for cancer onset, we assessed the association between their eigenvalues and cancer development using three complementary approaches: Cox proportional hazards models, Kaplan–Meier survival analysis, and logistic regression. Analyses were conducted in the EPIC-Italy cohort [65], a large prospective study in which blood DNA methylation was profiled at baseline in healthy participants, with up to 15 years of follow-up for incident cancer diagnoses. The cohort includes time-to-diagnosis information for several cancer types, with breast cancer (C50) and colorectal cancer (C18) being the most prevalent. We first assessed associations across all cancer types. Stratifying by sex revealed sex-specific predictive capacity. In females, NL3 showed the most consistent association with cancer risk across analyses. In Cox regression, NL3 eigenvalues were significantly associated with a shorter time to cancer diagnosis (HR 1.020, 95% CI: 1.007–1.032, FDR = 0.0058) (Additional file 5: Table S4). Kaplan–Meier curves showed a separation across NL3 tertiles (log-rank $p = 0.003$, Fig. 3D), and logistic regression yielded a concordant association (OR = 1.032, 95% CI: 1.012–1.052, FDR = 0.0083). NL11 also showed evidence of association. In the full female cohort, its effect was borderline in Cox regression (HR = 1.039, FDR = 0.063) but significant in logistic regression (OR = 1.065, 95% CI:

1.008–1.126, FDR = 0.049), suggesting a possible but weaker signal. No associations were detected for NL2 or NL12. In males, none of the clusters showed statistically significant associations. Next, we stratified the cohort by cancer type to investigate cancer-specific associations. In the breast cancer sub-cohort (female participants only), NL3 remained significantly associated with disease onset in both models (Cox HR = 1.023, 95% CI: 1.009–1.037, FDR = 0.0045; Logistic OR = 1.035, 95% CI: 1.013–1.058, FDR = 0.0051) (Additional file 5: Table S4). Likewise, NL11 showed a consistent association (Cox HR = 1.065, 95% CI: 1.019–1.113, FDR = 0.0097; Logistic OR = 1.101, 95% CI: 1.034–1.172, FDR = 0.0051). In contrast, we did not observe statistically significant associations between any nonlinear cluster and colorectal cancer (C18) in either males or females. Both Cox and logistic regression models yielded non-significant associations across all clusters, indicating that the epigenetic trajectories captured by these modules do not predict colorectal cancer onset within this cohort.

Inflammation—CRP levels

Senescence of the immune system is termed "inflammaging", with chronic inflammation being one of the hallmarks of aging [66, 67]. Particularly relevant to our blood-based analysis, we decided to examine the association between the nonlinear clusters and inflammation. We explored whether some of the age-related clusters we identified shared this predictive capacity. First, we looked at inflammation. Separately in males and females, we calculated each cluster's eigenvalues and looked at their association with estimated C-reactive protein (CRP) levels (Methods), a well-established marker of inflammation [68]. Three nested models were used to assess these associations while adjusting for age and NC CpGs (Methods). The analysis revealed sex-specific patterns of association, with several eigenvalues significantly predicting CRP independently of chronological age. In females, the full linear model including eigenvalues from all age-related clusters significantly improved CRP prediction compared to age + NC alone (adjusted $R^2 = 0.422$; ANOVA $p < 8.2 \times 10^{-11}$). Several modules showed robust associations (Fig. 3C, Additional file 5: Table S4). The NL3 module displayed the strongest positive association with CRP ($\beta = 0.0221$, $p = 2.06 \times 10^{-6}$), suggesting that methylation patterns in this cluster closely align with inflammatory status. The LD module was significantly negatively associated with CRP ($\beta = -0.0098$, $p = 1.34 \times 10^{-4}$), indicating a potential protective or anti-inflammatory methylation pattern. NL11 also revealed a significant negative association ($\beta = -0.0200$, $p = 0.023$), while VI showed a significant positive association ($\beta = 0.0179$, $p = 6.9 \times 10^{-4}$). Notably, the NC eigenvalue was not significantly associated with CRP ($p = 0.266$), although it contained 87% of the CpGs used in CRP estimations (Additional file 1: Fig. S4D). This result likely stems from the fact that we accounted for 12 immune cells fractions in our model, including naïve and mature Tells which the CRP signature relies on [68], and further shows that the NL patterns captured methylation variation independently of the immune profile. Together, these results suggest that inflammation in females is selectively captured by distinct age-related methylation patterns, some of which are positively associated with inflammatory burden (e.g., VI, NL3) and others inversely associated (e.g., LD, NL11), reflecting complex and potentially compensatory epigenetic dynamics. In contrast, males exhibited a more restricted profile of significant associations. While the full model remained statistically significant

(*adjusted* $R^2 = 0.423$; ANOVA $p = 4.82 \times 10^{-5}$), only the LI module showed a significant positive association with estimated CRP levels ($\beta = 0.0059$, $p = 0.0037$) (Fig. 3C, Additional file 5: Table S4). Associations with other clusters, such as NL1, NL2, and LD, did not reach significance, although some trends were observed. The NC module was again not predictive ($p = 0.65$), reinforcing the specificity of the signal to age-related modules. These sex-specific patterns suggest that while methylation-based inflammation signatures exist in both sexes, females display a more modular and interpretable architecture. In contrast, the inflammatory signal in males may be either more diffuse across the methylome or less tightly coupled to the cluster structure identified in the present analysis. Together, these findings indicate that methylation clusters derived from nonlinear age-related trajectories carry relevant information about the inflammatory burden in blood.

Different waves of regulation in males and females

In the analysis above, we identified functional clusters of age-related CpGs sharing similar linear or nonlinear patterns. Although this analysis hinted at specific ages at which disruption of methylation trajectories occurs (i.e., inflection points), another analytical method has been previously used to address this question directly [1, 6]. We applied the same method used by Shen & al [1]. to reveal peaks of disruption of methylation patterns in our male and female populations (Methods). Our analysis focused on age-related CpGs identified at the last step and revealed distinct waves of dysregulation in males and females (Fig. 4A left). Those peaks were deemed robust as they passed different significance thresholds (Fig. 4A right). In females, we observed three peaks at age 33, 51, and 73 that were consistent with 2 of the inflection points observed in the previous analysis. Whereas males only displayed two peaks at 47 and 63 years old, again consistent with the clustering analysis. The number of significantly dysregulated CpGs markedly differed between males and females, with females showing an overall higher number, similar to the discrepancy observed between their respective numbers of NL CpGs. Notably, the peaks at 33 and 51 in women are supported by previous findings highlighting those decades as transitional periods in women's aging [69] and previous research in protein expression levels identified three peaks at 34, 60, and 78 years old [6]. Recent findings in a multi-omics setting also identified crests of dysregulation around 40 and 60 years old. Both those analyses were performed in a joint cohort of men and women and might represent a combination of the sex-specific peaks we identified. Thus, our results support previous claims that aging occurs in waves across multiple modalities rather than unique molecular components. In addition, we highlighted a distinct temporality in the aging process across sex, as underlined by the earlier onset of the crests in males. Integrating this analysis with the clusters we identified previously revealed that the main CpG overlaps in males and females occur with the linear clusters LI and LD (Additional file 1: Fig. S5A-D). Most of the NL CpGs in females intersected with wave 3, whereas in males, they were equally represented in waves 1 and 2. This analysis further highlighted conserved CpGs across waves (N_males = 2027 & N_females = 1453) and across sexes ($N = 725$). Motif enrichment analysis of those sets confirmed the enrichment of NF1 full- and half-binding motifs at age-related CpGs, and revealed the presence of CTCF and NRSF/REST binding sites, two consistent marks of epigenetic aging [10, 34, 70–73],

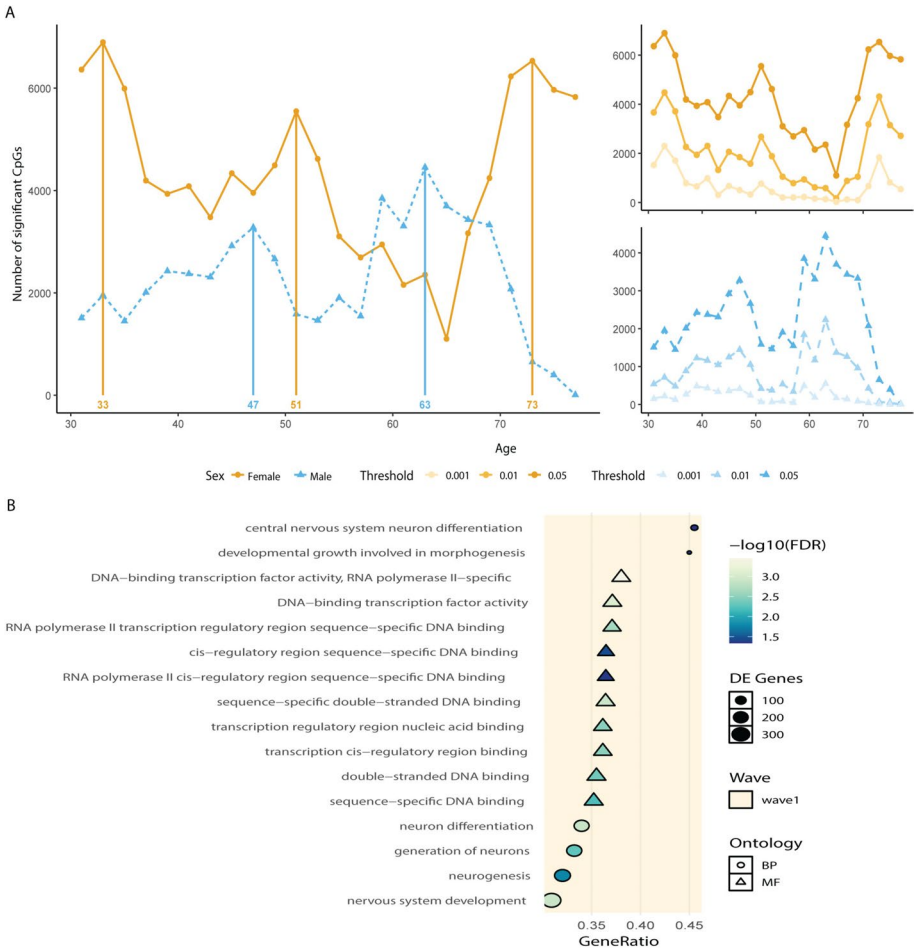

Fig. 4 **A** Left: DEswan analysis in males and females. Each point represents the number of CpGs dysregulated between a 15-year window on each side of the specified age at FDR < 0.05. Right: DEswan analysis at different FDR thresholds. **B** Pathway enrichment analysis on the CpGs at the three peaks identified in females. Only wave 1 showed significant term enrichment

among other TFBs (Additional file 1: Fig. S6A-D). To understand the specific pathways that could be affected, we repeated our previous enrichment analysis, focusing on each peak's unique set of CpGs (Methods). In females, waves 1 showed an enrichment for neurodevelopmental processes, including *nervous system development* (FDR = 0.0012), *neuron differentiation* (FDR = 0.0013), and *generation of neurons* (FDR = 0.005) (Fig. 4B). Additionally, terms related to transcriptional control and chromatin binding, such as *DNA-binding transcription factor activity* (FDR = 0.0008), and *double-stranded DNA binding* (FDR = 0.004) were significantly enriched (Additional file 6: Table S5). These findings recapitulate previous observations of neurodevelopmental and transcriptional gene dysregulation with age detected in blood [10, 45, 74, 75]. In males, among the unique CpGs in wave 1 ($N = 1250$) and wave 2 ($N = 2426$), no significant enrichment was found. Likewise, the 725 CpGs common to both sexes showed no detectable enrichment. This highlights sex-specific epigenetic changes in females, with broader and more functionally coherent methylation changes. Notably, REST, a transcriptional repressor enriched in our motif analysis, is known to silence neuronal genes in non-neuronal

lineages and is a key regulator of the epigenetic program that maintains cellular identity [73, 76]. Its involvement, alongside the dysregulation of neurodevelopmental pathways, supports a model in which age-related methylation drift compromises the fidelity of cell-type-specific gene regulation, allowing partial re-expression of lineage-inappropriate developmental programs. This aligns with emerging evidence from aging transcriptomes and methylomes indicating that immune cells in older individuals exhibit loss of identity and ectopic activation of developmental gene networks, including those tied to nervous system formation [77, 78].

Stability of methylation aging patterns

To assess the robustness and generalisability of the aging trajectories identified in our first cohort, we repeated our analysis in an additional cohort (GSE87571- Cohort 2). This dataset was generated on the Illumina 450 k array and includes samples for 385 females [14–94 yo] and 339 males [15–87 yo] (Additional file 1: Fig. S7A-B). Following pre-processing, we applied the SNITCH pipeline on the remaining autosomal CpGs ($n = 375{,}010$) while adjusting for estimated immune-cell composition (Methods). The resulting small number of CpGs associated with age (9% in females, 7% in males) is in line with our findings in the discovery cohort, but likely doesn't reflect the full amount of age-related changes, due to a low statistical power [79]. We found that the majority of the NL CpGs identified in the discovery cohort are associated with age in the new cohort (Males = 92%, Females = 78.9%), with only some classified as NL (F = 11.3%, M = 48.3%) (Additional file 1: Fig. S8A-B). Thus, although the effect of age on those CpGs seems to be stable, the exact pattern may vary depending on the population. Nevertheless, we found that the overall trajectory structure and dominant NL patterns were preserved (Fig. 5A, Additional file 1: Fig. S9D-10D), indicating that the main nonlinear aging signatures are consistent across those two populations. Notably, clusters with high correlation between the two cohorts shared a certain number of CpGs (Additional file 1: Fig. S11A-B, Additional file 7: Table S6). We found an enrichment for similar TF binding sites, recapitulating our findings in the first cohort (Additional file 1: Fig. S11C-D, Additional file 7: Table S6). In contrast to Cohort 1, we found an enrichment for certain pathways both in the female (M01) and male (M02) clusters with terms such as *transcription* and *development* identified (Additional file 1: Fig. S12A-B). No terms were significantly enriched for CpGs sharing the NL classification across the two cohorts. Finally, we repeated the Deswan analysis to assess the stability of the peaks of dysregulation identified in Cohort 1. This revealed peaks of dysregulation around age 31, 51 and 59yo in females, and 31 and 55 yo in males (Additional file 1: Fig. S13A), with the second wave in males enriched for pathways related to transcription and development (Additional file 1: Fig. S13B-C, Additional file 7: Table S6). Although the earlier peaks in females around the 3rd and 5th decades, and the 2nd peak in males, were recapitulated, both their size and the presence of other peaks contrasted across cohorts (Fig. 5B). Looking at the overlap between unique CpGs per wave across cohorts further underlined this disparity, as CpGs dysregulated in one wave in a cohort were often not represented in the corresponding wave of the second cohort (Fig. 5C-D).

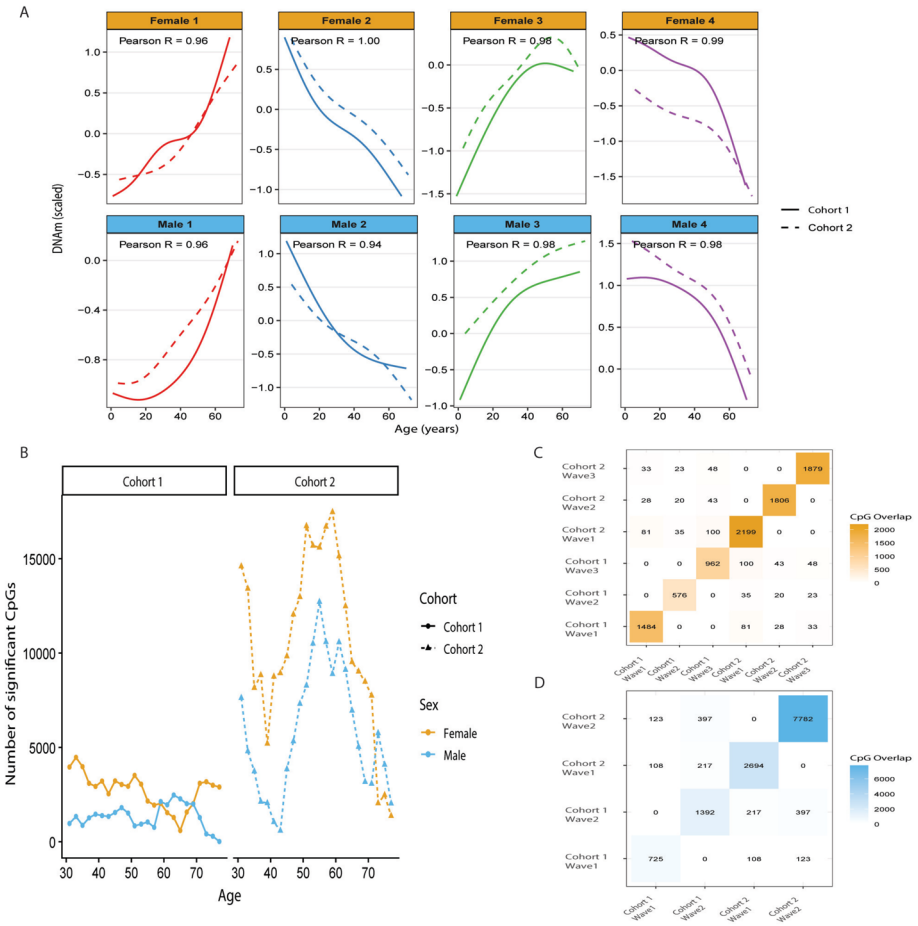

Fig. 5 **A** Conserved mean nonlinear aging trajectories in males and females across Cohorts 1 and 2. The Pearson correlation coefficient is indicated. **B** Results of the DEswan analysis in males and females across Cohorts 1 and 2. **C** Overlap between the unique sets of CpGs identified per wave in the two cohorts in females. **D** Overlap between the unique sets of CpGs identified per wave in the two cohorts in males. *Note:* The presence of zero in the matrices is due to the analysis being restricted to unique sets of CpGs per wave

Discussion

By leveraging a novel analysis pipeline and a high-quality DNA methylation (DNAm) dataset, we identified sex-specific nonlinear aging trajectories in blood. Fine-tuning our clustering pipeline on simulated data enabled us to establish robust heuristics facilitating the identification of diverging linear and nonlinear patterns (Fig. 1A, B). We showed that our new pipeline, SNITCH, outperforms stand-alone unsupervised clustering methods in discriminating between Variance Increase (VI), Linear Increasing and Decreasing (LI & LD), Non-Correlated (NC), and various Nonlinear (NL) functions (Fig. 1C, D). Nevertheless, we observed misclassification between the linear and logarithmic patterns due to their close resemblance (Fig. 1B, E). By providing SNITCH as a user-friendly framework, our approach is well-positioned to uncover nonlinear, tissue- or context-specific aging dynamics in large cross-sectional or longitudinal datasets, including EWAS, transcriptomic time series, or multi-omic aging studies.

The immune profile reshapes with age, and analyses in whole blood are particularly sensitive to those changes [80]. Applying our pipeline to a blood DNA methylation

dataset while accounting for an increasing number of immune cell types highlighted the confounding effect of the immune profile on age-related DNAm changes (Fig. 2A). Stratifying our analyses across sexes revealed a notably higher number of age-related and NL CpGs identified in Females in both cohorts (Fig. 2B, Additional file 1: Fig. S8A), agreeing with previous findings [81]. Nevertheless, CpGs generally showed concordant aging trajectories across sexes, with significant enrichment for matching classifications (Additional file 1: Fig. S1C). The small number of age-related CpGs identified (< 4% in cohort 1 and < 10% in cohort 2) seemingly contrasts with findings of a meta-analysis of age-related DNAm in blood [18]. The meta-analysis provides sensitivity to detect CpGs with small effect sizes that are likely below the detection threshold in our current cohorts, while our stricter QC and the inclusion of 12 immune cell fractions likely also contribute to our more conservative results.

Our study assessed the distribution of clock CpGs within our different aging clusters and found that most models relied on CpGs classified as NC in our model (Fig. 3A). Although counterintuitive, we focused on changes independent of the immune profile, where clocks are known to be susceptible to cell fraction changes [82]. This was supported by our baseline model, which showed fewer CpGs classified as non-correlated. Additional explanation comes from the data used to train the clocks. Indeed, among the 9 models we investigated, only the Hannum clock was trained solely on blood DNAm using chronological age as the target variable, where the other clocks were trained on multiple tissues (Horvath's) or phenotypic variables (PhenoAge, GrimAge, DunedinPace, Zhang_10, and YingAge). We found that most of the clocks included VI and NL CpGs. Further analyses should investigate the weight associated with those in each model to understand the reliance of the clocks on nonlinear trajectories, and potentially fine-tune them to reflect the nonlinear phases of aging. However, the CpGs analysed do not capture the full range of sites associated with biological aging, but instead represent a set of model-specific features [83].

Chromatin state enrichments aligned with what is known in the literature [18] irrespective of sex or linearity (Fig. 3B). Only the NL3 cluster in females differed from this trend and showed a small enrichment for the *weak transcription* state. Consistent with other results, the motif enrichment analysis revealed an enrichment in the motif of REST-NRSF [72] in the male hypermethylated NL1 cluster. The loss of REST is associated with cognitive impairment and Alzheimer's disease [73], and hypermethylation at its binding sites could disrupt its function. Although previous studies have highlighted blood DNAm marks as biomarkers for neurological phenotypes [74, 84, 85], the relevance of this finding in blood should be further explored. This analysis also identified new enriched motifs at NL CpGs for the NF1-CTF family, both in males and females, suggesting a potential disruption of their function. Additional transcription factors (TFs) having their binding sites enriched included the STAT family in males (STAT3 & STAT4), underlying a potential disruption of the immune system [86] and developmental TFs in females (HOXC9, GATA6) [55, 56]. Notably, the enrichment of motifs from developmental TFs at hypomethylated sites could indicate the loss of a safeguarding mechanism consistent with the epigenetic drift view. Nevertheless, our analyses suggest that this drift is not constant but suffers from episodes of dysregulation.

The functional role of our NL clusters was further explored by looking at their association with inflammation and cancer onset in an independent cohort. Accounting for immune cell fractions, we identified NL3 in females as being associated with both inflammation (Fig. 3D) and cancer onset (Fig. 3C), and the female NL11 cluster as mildly significant. The functional specificity of those clusters was underscored by the absence of predictive power from the NC module across sexes. Furthermore, the more robust and multifaceted associations in females support the argument for sex-stratified biomarkers, particularly in the context of aging, cancer, and inflammation. Although the associations between blood-derived DNAm biomarkers and cancer onset have been documented [63, 64], the underlying mechanisms remain unclear. A potential link would be through immune infiltration [87]. Another possibility is that DNAm captures systemic effects linked to age, where changes in blood would mirror those in the tissue of interest [37]. Our motif enrichment analysis laid the ground for further investigating these mechanisms in females NL3. Notably, NF1/CTF, Hoxc9, and Gata6 binding motifs were overrepresented in this cluster. Gata6 was recently shown to be part of a central mechanism promoting cancer-associated fibroblasts in breast cancer [58]. The study found that the expression of Gata6 was enhanced by the action of TET1, a protein that removes methylation. Evidence shows that binding of members of the Gata family is disrupted by methylation marks [88]. Thus, the hypomethylation observed in NL3 could be associated with increased Gata6 oncogenic activity. Both Hoxc9 and members of the NF1 family have also been associated with breast cancer, although the mechanisms have not been fully elucidated and are not well documented [50, 57, 89]. Nevertheless, how the occurrence of these epigenetic marks in immune cells may affect cancer onset in other tissues remains to be investigated.

Overall, our clustering analysis revealed sex-specific functional clusters of NL CpGs. They hinted at sex-specific peaks of dysregulation by showing inflection points in mid and late life, agreeing in both cohorts (Fig. 5A). We used the previously described DEswan analysis [1] to formally identify peaks of dysregulation in a sex-specific manner. The DEswan results reinforced the findings of our first analysis by revealing peaks of dysregulation in the first cohort at 33, 51 and 73 years old in females, and 47 and 63 years old in males (Fig. 4A). In a second cohort with identified peaks at 31, 51 and 59 in females and 31 and 55 in males. This hints at conserved periods of dysregulation with studies in proteomics and multi-omics identifying peaks at similar ages [1, 6, 69], but also points toward population specificity. The relatively small size of our cohorts warrants caution in a definitive conclusion regarding the exact timepoint of those peaks and the minimum age of 25 in our cohort, probably missing prior peaks during adolescence [90]. Motifs enrichment for sex-specific and non-sex-specific conserved CpGs supported our previous findings with NF1/CTF and REST amongst the most enriched binding sites. Consistent with other findings [71, 72] but absent from our cluster analysis, the CTCF motif was also enriched. Our functional analysis for the waves of dysregulation highlighted contrasting results across cohorts with little sex- or time-specificity in pathway enrichment. Nevertheless, significant pathways associated with nonlinearity consistently pointed toward regulation of transcription and development (Fig. 4B, Additional file 1: Figs. S12-13), in agreement with previous findings [45, 74, 75, 91].

Our study presents several limitations. First, the datasets lacked critical covariates such as BMI or smoking status, preventing us from fully disentangling lifestyle effects from intrinsic aging-related methylation changes. This could be addressed by tools predicting lifestyle habits [92, 93]. Additionally, both datasets were predominantly composed of individuals of European ancestry, which may limit the generalizability of our findings to more diverse populations, although a study in a Bangladeshi population recapitulates the findings that females tend to show more changes in DNAm with age [81]. Despite covering a wide age range (18–90 years old), we missed samples in the early infancy/adolescent stages, known to be associated with wide changes in DNAm [90, 94, 95]. As a result, our characterization of methylation dynamics during early development remains incomplete, potentially missing early-life inflection points critical to the trajectory of aging. Another limitation is the cross-sectional nature of our study. This is a typical limitation in the field due to the lack of large longitudinal datasets. Thus, our analysis relies on the hypothesis that DNAm changes are conserved to a level across different individuals, a hypothesis supported by the large body of literature on DNAm in aging. Nevertheless, adopting a longitudinal approach is an important step towards personalised medicine and is a direction the field should embrace. In this work, we analysed datasets produced at single centres to reduce batch effects. While this strategy still allowed detection of conserved changes validated in an independent cohort, our relatively small sample sizes ($N < 500$) likely limited our ability to detect subtle but biologically relevant effects [79], reflected by the small number of age-related CpGs we found. Thus, our results should be replicated across larger cohorts. Finally, DNAm is intrinsically difficult to functionally link to changes in the phenotype, in part due to the variety of effects that gain or loss of methylation can have depending on their location. Consequently, we emphasize that our findings are correlative and should be interpreted as surrogate markers rather than direct drivers of phenotypic changes.

In the future, we aim to further explore nonlinear patterns in other tissues [96] and extend our methodology to other omics, such as transcriptomics and proteomics, where the increase of variance is often omitted from clustering analyses. Our results, along with prior work, reveal enrichment of developmental transcription factor motifs (e.g., REST, NF1/CTF) at age-dysregulated CpGs [37, 47], supported by the enrichment of pathways associated with transcription and development. This observation remains underexplored, particularly regarding how such TF-DNAm interactions shift with age, impact transcriptional regulation, and affect cellular identity. Future studies integrating these findings with transcriptomic or chromatin accessibility data could elucidate the functional consequences of these methylation changes.

Conclusions

Altogether, our results are consistent with aging being a nonlinear, entropy-increasing process characterized by discrete windows of heightened epigenetic instability, rather than a simple linear decline. By uncovering sex-specific, wave-like patterns of DNA methylation changes that mirror inflection points reported in other omics layers, our study strengthens the case for integrated, temporal frameworks of aging biology. The critical decades we identified, particularly around the 30 s, 50 s, and 70 s, may represent biologically vulnerable windows, where intervention could yield the most impact.

Beyond these insights, our findings underscore the need to move beyond one-size-fits-all models and adopt analytical strategies that reflect the inherent heterogeneity, nonlinearity, and sex-specific nature of biological aging. However, to fully harness the translational potential of these observations, a pressing need for molecular validation remains. Future work should prioritize characterising the regulatory mechanisms underpinning these methylation dynamics, including their interaction with chromatin architecture, transcription factor networks, and other epigenetic layers, to distinguish causality from correlation and to illuminate actionable pathways of aging and disease.

Methods

Cohort and DNA methylation preprocessing

Three separate cohorts were used in this study. GSE246337 [97] and GSE87571 [75] were used in the identification of aging methylation patterns, and GSE51032 [98] was used to identify biomarkers of cancer. For all datasets, the Raw IDAT files from the Illumina Infinium HumanMethylationEPIC v2 BeadChip array (EPICv2) or 450 K, were retrieved from GEO and processed using the *sesame* R/Bioconductor package (v1.18.1) [99]. To harmonize probe identifiers and accommodate the EPICv2 platform structure, prefix collapsing was enabled during data import. Initial preprocessing followed the "QCDPB" pipeline within *sesame*, incorporating: (i) probe quality filtering using the pOOBAH method to remove probes with poor detection *p*-values, (ii) dye-bias correction to normalize type I and II probe discrepancies, (iii) masking of probes known to be problematic due to non-specific hybridization or SNP interference as defined by Zhou et al. [100], and (iv) exclusion of probes supported by fewer than four beads. The resulting beta value matrix was further filtered to improve data integrity. Probes with missing values in more than 1% of samples were discarded. Subsequently, samples with more than 1% missing beta values across retained CpGs were also removed, and we removed probes on the sex chromosomes. The remaining missing values were imputed using k-nearest neighbors using the impute.knn functions from the *impute* package, keeping the default parameters. Beta values were rounded to three decimals for downstream analyses. For the GSE246337 dataset, we only kept probes present on the EPICv1 array, as most tools used in the downstream analysis are not yet compatible with the EPICv2 array. The final number of probes used in the downstream analyses was 556,811, and the final numbers of samples were 256 females and 238 males for GSE246337, 385 females, 339 males and 375,010 probes for GSE87571. For the GSE51032 [98] cohort, the final number of probes was 382,717, for 651 females and 186 males.

Description of the SNITCH analysis pipeline

The SNITCH pipeline involved three main steps: Heuristic-based classification, Functional Principal Components Analysis (FPCA), and unsupervised clustering.

Heuristic-based classification

For each CpG site, methylation beta values are modeled as a function of chronological age. The core procedure included the following steps:

1. Model Construction: Linear models (LMs) are fitted using ordinary least squares regression with age as a continuous predictor (lm function from base R). Parallelly, generalized additive models (GAMs) are fitted using restricted maximum likelihood (REML) and thin plate regression splines (*s(Age, k=5)*), enabling the detection of nonlinear trends (gam function - *mgcv* [101]). When relevant, covariates are included consistently across both models to preserve interpretability and comparability.
2. Model Comparison and Heteroscedasticity Testing: The explanatory performance of GAMs relative to LMs is evaluated using the Bayesian Information Criterion (BIC) (BIC function from base R). CpGs with ΔBIC (BIC(LM) − BIC(GAM)) > 2 are considered to favor the nonlinear model. To characterize potential violations of homoscedasticity that may underlie complex aging dynamics (VMP), White's test is applied to each LM, with heteroscedastic CpGs flagged based on a 1% FDR-adjusted significance threshold.
3. Prediction and Effect Size Estimation: For CpGs favoring a GAM fit, DNAm values are predicted across a continuous age grid ranging from the minimum age of the cohort to the maximum with a one-year step increase, holding covariates constant at reference levels (medians for numeric or first level for categorical variables). This effectively smoothed the trajectories for subsequent steps. For linear trajectories, the direction and significance of age-associated change are inferred from the LM coefficient and *p*-value, respectively.
4. Multiple Testing and Classification: *P*-values from LM, GAM, and heteroscedasticity tests are corrected using the Benjamini-Hochberg method. CpGs are classified as:

    - LI (Linear Increase): Significant linear association (adj. p(LM) ≤ 0.01) with a positive slope.
    - LD (Linear Decrease): Significant linear association with a negative slope.
    - NL (Nonlinear): ΔBIC > 2 and adj. p(GAM) ≤ 0.01.
    - VI (Variance-Increasing): No linear association (adj. p(LM) > 0.01) but significant heteroscedasticity (adj. p(White) ≤ 0.01).
    - NC (Non-correlated): CpGs not meeting any of the above criteria.

This classification strategy allows SNITCH to robustly detect a spectrum of epigenetic aging signatures, from canonical linear changes to more complex nonlinear patterns. The SNITCH method is available and can be installed as a user-friendly R package (https://github.com/fishrscale/SNITCH), thus facilitating access to nonlinear trajectory analyses.

Functional principal component analysis (FPCA)

To further dissect heterogeneity within nonlinear (NL) CpG methylation trajectories, we implemented a two-stage dimensionality reduction and clustering procedure based on functional principal component analysis (FPCA) followed by density-based unsupervised classification. This enabled the grouping of NL CpGs into discrete functional subclusters based on the shape of their smoothed age-associated methylation trajectories.

For all CpG sites previously classified as NL by SNITCH, we use the smoothed beta values predicted by the GAM model. FPCA is then applied using the fpca.face() function from the *refund* R package, with age as the functional domain. The number of knots is set dynamically based on the number of time points (min(35, floor(0.8 * timepoints))), and the proportion of variance explained (PVE) threshold is conservatively fixed at 99.99% to retain fine-grained trajectory information. This process decomposes the complex, high-dimensional nonlinear patterns into a reduced set of orthogonal functional basis scores.

Unsupervised clustering

The resulting FPCA scores are used as input for unsupervised clustering using either the hdbscan (*dbscan*), fuzzy-clustering (*mfuzz)*, or kmeans (base R) algorithm. This procedure allowed data-driven identification of methylation trajectory subtypes without requiring prior knowledge of cluster number or shape.

Simulated data

A total of 3,000 synthetic CpG sites were simulated across 300 individuals, each assigned a random age between 1 and 100 years. Fifteen trajectory archetypes were implemented to represent a range of biologically plausible methylation patterns. These included: non-correlated, linear trajectories (increasing, decreasing), quadratic trends (increasing, decreasing), logarithmic transitions (increasing, decreasing), sigmoidal dynamics (increasing, decreasing) with inflection points at ages 25, 40, and 80, variance-increasing profiles, and non-monotonic patterns. For each of the 15 trajectory classes, 200 CpGs were simulated. For all the functions except the variance-increasing, the age-specific methylation expectation (μ) was passed to a Beta distribution (via rbeta) to introduce variability while maintaining biological constraints (bounded between 0 and 1). For the variance-increasing function, age-dependent Gaussian noise was added directly to a mean of 0.5, with standard deviation increasing linearly from 0.01 (age < 25) to 0.3 (age = 100), mimicking stochastic methylation drift.

Benchmarking SNITCH

The benchmarking was done using the simulated patterns and their associated labels as ground truth. Fuzzy c-means clustering was performed using the *Mfuzz* package. The fuzzification parameter *m* was estimated using the function mestimate, and the optimal number of clusters was set to 11 after determining the minimum centroid distance across 3 repeated runs. Final cluster assignments were defined by the highest membership value. K-means was run with 25 restarts and centers = 10 by using the elbow method based on total within-cluster sum of squares (WCSS) and k.max = 20. HDBSCAN was applied using the *dbscan* package, the minimum cluster size was set to 5. The DICNAP pipeline was implemented as originally described [19], except for the maximal number of clusters for K-means set to 20. SNITCH was run as previously described. The

FPCA scores were subsequently used for unsupervised classification by the same three algorithms. The '*c*' parameter was set to 11 for fuzzy clustering, and centers = 10 for K-means. Each method's cluster assignments were compared to ground-truth labels to compute ARI and AMI using the same functions from *aricode*. The results from the two best-performing methods were evaluated by a confusion matrix.

Identification of sex-specific CpG Methylation Trajectories Using SNITCH

GSE246337

To evaluate aging-associated methylation trajectories independently of immune cell composition, we constructed three models differing in their treatment of cellular heterogeneity. In the Baseline (BL) model, we ran SNITCH with default parameters and without correcting for covariates. In the 7-cell and 12-cell models, we accounted for cell type proportions estimated using the *EpiDISH* package with the centDHSbloodDMC.m and cent12CT.m reference matrix, respectively, using the Robust Partial Correlations (RPC) method [25]. The 7 cell–matrix contained information on B-, NK-, CD4T, and CD8T-cells, Monocytes, Neutrophils, and Eosinophils. Building on the 7-cell reference matrix, the 12-cell model discriminated between naive and mature B-, CD4T-, and CD8T-cells, and added T-regulatory cells and Basophils. To identify groups of CpGs with similar aging dynamics, we applied both *Mfuzz* and *HDBSCAN* to the SNITCH-classified nonlinear trajectories. Clustering results were assessed visually, and we retained only the *HDBSCAN* clusters for downstream analyses due to their superior trajectory homogeneity. Similarly, the optimal minPts parameter for HDBSCAN was determined by visual inspection of cluster resolution. The selected minPts values for the female BL, 7-cell, and 12-cell models were 5, 5, and 7, respectively. For the male models, they were 5, 8, and 4. Notably, HDBSCAN designates sparse or noisy data points as cluster 0. Upon reviewing the trajectories assigned to this group in the female 12-cell model, we observed two distinct patterns within cluster NL1. We therefore re-ran HDBSCAN on this subset to refine the classification, resulting in three clusters: NL10, NL11, and NL12. Then, for each NL cluster, we performed principal component analysis (PCA) and extracted PC1 to represent the dominant methylation pattern. A correlation matrix was then computed across all cluster PC1 using Spearman's rank correlation. Clusters with correlations exceeding 0.90 were merged to reduce redundancy and capture shared underlying dynamics (Additional file 1: Figs. S2-3).

GSE87571

We repeated the steps above for the female and male samples within the GSE87571 cohort. Notably, we only implemented the 12-cell model, using the cent12CT450k.m reference matrix in EpiDISH, and we used the *Mfuzz* package for the classification of nonlinear CpGs with c_opt set to 10. Similarly, we merged clusters with a Spearman correlation above 0.9 (Additional file 1: Figs. S9-10).

Functional analysis

Classification of clocks and immune cells CpGs

CpGs were retrieved for the following clocks from the Biolearn resource [97]: Horvath v1, Hannum, PhenoAge, GrimAgeV2, DunedinPACE, YingCausAge, YingDamAge, YingAdaptAge, and Zhang_10. Unique CpG identifiers were retained. We counted the occurrence of each clock CpGs in the different categories identified by SNITCH across our 3 models (BL, 7-cell, 12-cell) in both males and females. A similar analysis was performed with age-related CpGs (169 hypermethylated, 181 hypomethylated) shared across immune cells retrieved from Roy R. et al. [34].

Chromatin state enrichment

Chromatin states profiled in peripheral blood mononuclear cells (PBMCs) were obtained from the Roadmap Epigenomics Project [35]and mapped to CpG probe IDs. Fisher's exact tests were conducted to evaluate the enrichment of each chromatin state across aging trajectory classes against NC CpGs in both sexes. *P*-values were adjusted using the Benjamini–Hochberg method.

Pathway enrichment analysis

The gometh function from the *missMethyl* [44]R package was used to perform Over Representation Analysis (ORA) in the Gene Ontology [102] and KEGG [103]databases. This function accounts for the differing numbers of CpG probes per gene on the Illumina EPIC array, thereby reducing bias inherent to standard enrichment tools when applied to DNAm data. We used the set of CpG sites corresponding to each DNA methylation cluster, excluding the non-correlated "NC" sites, as the test set, and the full set of tested CpG sites across all clusters served as the background. Analyses were conducted independently for male and female clusters. The analysis was replicated for the CpGs of each peak identified during the DEswan analysis. The analysis was replicated for the CpGs of each peak identified during the DEswan analysis. When relevant, the *simplifyEnrichment* package [104] was used to group similar terms, using default parameters.

Motif enrichment analysis

The motif enrichment tool from EWAS datahub [105] was used to perform motif enrichment analysis on each cluster identified in females and males. This tool uses a centered 500 bp window on the CpG of interest to perform Hypergeometric Optimization of Motif EnRichment (HOMER) [106]. Motifs with a qvalue < 0.05 were considered significant. The analysis was replicated for the CpGs of each peak identified during the DEswan analysis.

Trait association—cancer

The cohort used to assess the sex-specific biomarker potential of cluster eigenvalues has been described elsewhere. Briefly, the EPIC-Italy [65] study includes DNA methylation data collected at baseline and linked to up to 14 years of prospective follow-up. Cancer

cases were annotated with time to diagnosis and cancer type, enabling time-to-event analyses. We performed sex-stratified analyses using DNAm of patients preprocessed as described above. Individuals without a cancer event were treated as right-censored. For these, the time to diagnosis was imputed using the maximum observed follow-up time among cases. For each patient, we computed their NL cluster eigenvalues as described above. Importantly, this step was restricted to the CpGs common to both the EPIC and 450 K arrays, resulting in smaller sets of CpGs per cluster. Immune cell-type proportions were estimated from whole blood methylation profiles using the EpiDISH [25]algorithm (RPC method) with a 12-cell reference panel. The coxph function from the *survival* package was used to fit cox proportional hazards models using "Surv(time_to_diagnosis, status)" as the outcome, with eigenvalues, age, and immune cell proportions as predictors. Logistic regression models (glm(family = binary)) using cancer status as the binary outcome were also fit for comparison. Model coefficients were exponentiated to yield hazard ratios (HR) and odds ratios (OR), with 95% confidence intervals. To assess survival differences, samples were stratified into tertiles based on the significant eigenvalues (Low, Mid, High). Kaplan–Meier curves were generated using the survfit function and plotted with ggsurvplot() (log-rank p-value and 95% CI shown) from the *survminer* package. Significance was considered as FDR < 0.05.

Trait association—inflammation

Separately in males and females, we computed the "eigenvalue" of each cluster, corresponding to that cluster's first principal component (PC1), by performing principal component analysis (PCA) on centered and scaled methylation beta values across CpGs within that cluster. For each sample, we used the *ComputeCRPscore* [68] function based on a predefined set of weighted CpGs to generate an estimation of CRP protein levels as a surrogate for inflammation [107]. To test the independent contribution of age-associated methylation modules to CRP variation, we constructed a series of nested linear models:

- Model 1 included age as the only predictor;
  CRP ~ age
- Model 2 included age and the eigenvalue of the NC (non-correlated) CpGs;
  CRP ~ age + NC
- Model 3 (full model) included age, the NC eigenvalue, and eigenvalues of the VI, LI, LD, and NL clusters.
  CRP ~ age + NC + VI + LI + LD + NLi + … + NLn

We compared models using ANOVA to evaluate whether the addition of SNITCH-derived modules significantly improved the explanation of CRP variability beyond age and (non-correlated) methylation patterns. Model fit, and the significance of individual predictors were assessed using standard linear modeling statistics and p-values < 0.05.

Wave of aging—DEswan analysis

We performed a DEswan analysis as previously described [1, 6]. DEswan works as a sliding window, where at each specified time point centered at the middle of the window,

CpGs at both ends are compared using a Wilcoxon test for differential methylation. We restricted analyses to age-associated CpGs previously identified (excluding NC CpGs). We used windows centered from 31 to 78 years in 2-year steps with a 15-year bucket size. To account for cell composition, we estimated per-sample blood cell-type proportions with EpiDISH using a 12-cell reference (cent12CT for GSE246337and cent12CT450k.m for GSE87571) and included these estimates as covariates in all models (as described above). *P*-values were adjusted for multiple testing by the Benjamini–Hochberg method, and significance was considered at FRD < 0.05. To ensure the robustness of the findings, we only considered peaks that were conserved at FDR < 0.01 and 0.001 (Fig. 4B). We adapted the initial R script to allow parallelization across CpGs using parLapply.

Robustness of nonlinear CpGs

We computed the average trajectory per cluster in the two cohorts and used Pearson correlation to assess similarity.

Abbreviations

DNAm	DNA methylation
LI	Linear Increasing
LD	Linear Decreasing
VI	Variance Increasing
NC	Non-Correlated
NL	Nonlinear
ARI	Adjusted Rand index
AMI	Adjusted Mutual Information
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
TF	Transcription Factor
TFB	Transcription Factor Binding Site
CRP	C-Reactive Protein

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-026-03952-z.

Additional file 1. Supplementary Figures S1-13.

Additional file 2: Table S1. CpGs classification in males and females after the SNITCH pipeline.

Additional file 3: Table S2. Chromatin states enrichment analysis across SNITCH clusters.

Additional file 4: Table S3. Enrichment analysis of GO, KEGG terms, and TF motifs.

Additional file 5: Table S4. CRP and Cancer regression analyses.

Additional file 6: Table S5. Enrichment analyses across the waves of aging.

Additional file 7: Table S6. Enrichment analyses across SNITCH clusters and the waves of aging.

Acknowledgements

We would like to thank everyone who supports open-access data.

Peer review information

Weiqi Zhang and Claudia Feng were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions

RG performed data analysis and interpretation. RG, AT, and NE conceived and designed the study. AT provided methodological and computational modeling expertise. AT and SH provided guidance on DNA methylation analysis and interpretation of the results. RG & NE drafted the manuscript. NE, MJ and BJF contributed to ideas and revisions on the manuscript. All authors approved the final version of the manuscript.

Funding

RG is a recipient of the WBI Excellence World. This work was supported by Nir Eynon's NHMRC Investigator Grant (APP1194159), Hevolution/AFAR New Investigator Awards in Aging Biology and Geroscience Research, and the Australian Research Council (ARC) Discovery Project Grant (DP240102155).

Data availability
All data analysed during this study are publicly available in the GEO repository under accession numbers GSE246337 [97], GSE87571 [75], and GSE51032 [98]. The SNITCH method and scripts used in this study are accessible with open access at the following Github: (https:/github.com/fishrscale) [108] and Zenodo: (https:/zenodo.org/records/18203539) [109] repositories, licensed under the Apache License, Version 2.0.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 2 September 2025 Accepted: 12 January 2026
Published online: 04 February 2026

References

1. Shen X, et al. Nonlinear dynamics of multi-omics profiles during human aging. Nat Aging. 2024;4:1619–34.
2. Ye Q, et al. Telomere length and chronological age across the human lifespan: a systematic review and meta-analysis of 414 study samples including 743,019 individuals. Ageing Res Rev. 2023;90:102031.
3. Schaum N, et al. Ageing hallmarks exhibit organ-specific temporal signatures. Nature. 2020;583:596–602.
4. Kang Y-K, Min B, Eom J, Park JS. Different phases of aging in mouse old skeletal muscle. Aging. 2022;14:143–60.
5. Fehlmann T, et al. Common diseases alter the physiological age-related blood microRNA profile. Nat Commun. 2020;11:5958.
6. Lehallier B, et al. Undulating changes in human plasma proteome profiles across the lifespan. Nat Med. 2019;25:1843–50.
7. Aramillo Irizar P, et al. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. Nat Commun. 2018;9:327.
8. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018;19:371–84.
9. Seale K, Horvath S, Teschendorff A, Eynon N, Voisin S. Making sense of the ageing methylome. Nat Rev Genet. 2022;23:585–605.
10. Lu AT, et al. Universal DNA methylation age across mammalian tissues. Nat Aging. 2023;3:1144–66.
11. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14:3156.
12. Freire-Aradas A, et al. A common epigenetic clock from childhood to old age. Forensic Sci Int Genet. 2022;60:102743.
13. Johnson ND, et al. Non-linear patterns in age-related DNA methylation may reflect CD4+ T cell differentiation. Epigenetics. 2017;12:492–503.
14. Vershinina O, Bacalini MG, Zaikin A, Franceschi C, Ivanchenko M. Disentangling age-dependent DNA methylation: deterministic, stochastic, and nonlinear. Sci Rep. 2021;11:9201.
15. Olecka M, et al. Nonlinear DNA methylation trajectories in aging male mice. Nat Commun. 2024;15:3074.
16. Luo Q, et al. A meta-analysis of immune-cell fractions at high resolution reveals novel associations with common phenotypes and health outcomes. Genome Med. 2023;15:59.
17. Slieker RC, et al. Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. Genome Biol. 2016;17:191.
18. Seale K, Teschendorff A, Reiner AP, Voisin S, Eynon N. A comprehensive map of the aging blood methylome in humans. Genome Biol. 2024;25:240.
19. Okada D, Cheng JH, Zheng C, Kumaki T, Yamada R. Data-driven identification and classification of nonlinear aging patterns reveals the landscape of associations between DNA methylation and aging. Hum Genomics. 2023;17:8.
20. Phyo AZZ, et al. Sex differences in biological aging and the association with clinical measures in older adults. Geroscience. 2024;46:1775–88.
21. Sampathkumar NK, et al. Widespread sex dimorphism in aging and age-related diseases. Hum Genet. 2020;139:333–56.
22. Reicher L, et al. Phenome-wide associations of human aging uncover sex-specific dynamics. Nat Aging. 2024;4:1643–55.
23. Yusipov I, et al. Age-related DNA methylation changes are sex-specific: a comprehensive assessment. Aging. 2020;12:24057–80.
24. Qi L, Teschendorff AE. Cell-type heterogeneity: why we should adjust for it in epigenome and biomarker studies. Clin Epigenetics. 2022;14:31.
25. Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. BMC Bioinformatics. 2017;18:105.
26. Teschendorff AE, Horvath S. Epigenetic ageing clocks: statistical methods and emerging computational challenges. Nat Rev Genet. 2025. https://doi.org/10.1038/s41576-024-00807-w.

27. Belsky DW, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. Elife. 2022;11:e73420. https://doi.org/10.7554/eLife.73420.
28. Hannum G, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49:359–67.
29. Levine ME, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging. 2018;10:573–91.
30. Zhang Y, et al. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. Nat Commun. 2017;8:14617.
31. Lu AT, et al. DNA methylation GrimAge version 2. Aging. 2022;14:9484–549.
32. Ying K, et al. Causality-enriched epigenetic age uncouples damage and adaptation. Nat Aging. 2024;4:231–46.
33. Lu AT, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging. 2019;11:303–27.
34. Roy R, et al. Epigenetic signature of human immune aging in the GESTALT study. Elife. 2023;12:e86136.
35. Bernstein BE, et al. The NIH roadmap epigenomics mapping consortium. Nat Biotechnol. 2010;28:1045–8.
36. Moqri M, et al. PRC2-AgeIndex as a universal biomarker of aging and rejuvenation. Nat Commun. 2024;15:5956.
37. Teschendorff AE, 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.
38. Rakyan VK, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. 2010;20:434–9.
39. Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics Chromatin. 2018;11:25.
40. Jain N, et al. DNA methylation correlates of chronological age in diverse human tissue types. Epigenetics Chromatin. 2024;17:25.
41. Bartz J, Jung H, Wasiluk K, Zhang L, Dong X. Progress in discovering transcriptional noise in aging. Int J Mol Sci. 2023. https://doi.org/10.3390/ijms24043701.
42. Zhao K, Rhee SY. Interpreting omics data with pathway enrichment analysis. Trends Genet. 2023;39:308–19.
43. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. Genome Biol. 2021;22:173.
44. Phipson B, Maksimovic J, Oshlack A. Missmethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32:286–8.
45. Welsh H, et al. Age-related changes in DNA methylation in a sample of elderly Brazilians. Clin Epigenetics. 2025;17:17.
46. Rimoldi M, et al. DNA methylation patterns of transcription factor binding regions characterize their functional and evolutionary contexts. Genome Biol. 2024;25:146.
47. Yin Y, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017;356:eaaj2239.
48. Malaymar Pinar D, et al. Nuclear factor i family members are key transcription factors regulating gene expression. Mol Cell Proteomics. 2025;24:100890.
49. Chen K-S, Lim JWC, Richards LJ, Bunt J. The convergent roles of the nuclear factor I transcription factors in development and cancer. Cancer Lett. 2017;410:124–38.
50. Perumal N, et al. Nuclear factor i/B: duality in action in cancer pathophysiology. Cancer Lett. 2025;609:217349.
51. Gronostajski RM. Roles of the NFI/CTF gene family in transcription and development. Gene. 2000;249:31–45.
52. Liu Y, et al. PD-L1-mediated immune evasion in triple-negative breast cancer is linked to the loss of ZNF652. Cell Rep. 2023. https://doi.org/10.1016/j.celrep.2023.113343.
53. Kumar R, et al. ZNF652, a novel zinc finger protein, interacts with the putative breast tumor suppressor CBFA2T3 to repress transcription. Mol Cancer Res. 2006;4:655–65.
54. Kumar R, et al. CBFA2T3-ZNF652 corepressor complex regulates transcription of the E-box gene HEB*. J Biol Chem. 2008;283:19026–38.
55. Tiyaboonchai A, et al. GATA6 plays an important role in the induction of human definitive endoderm, development of the pancreas, and functionality of β pancreatic cells. Stem Cell Reports. 2017;8:589–604.
56. Wang X, et al. *HOXC9* directly regulates distinct sets of genes to coordinate diverse cellular processes during neuronal differentiation. BMC Genomics. 2013;14:830.
57. Hur H, et al. HOXC9 induces phenotypic switching between proliferation and invasion in breast cancer cells. J Cancer. 2016;7:768–73.
58. Ghazimoradi MH, Babashah S. The transcriptional regulators GATA6 and TET1 regulate the TGF-β pathway in cancer-associated fibroblasts to promote breast cancer progression. Cell Death Discov. 2025;11:164.
59. Giuili E, et al. Comprehensive evaluation of the implementation of episignatures for diagnosis of neurodevelopmental disorders (NDDs). Hum Genet. 2023;142:1721–35.
60. Grolaux R, et al. Identification of differentially methylated regions in rare diseases from a single-patient perspective. Clin Epigenetics. 2022;14:174.
61. Draškovič T, Hauptman N. Discovery of novel DNA methylation biomarker panels for the diagnosis and differentiation between common adenocarcinomas and their liver metastases. Sci Rep. 2024;14:3095.
62. Wang T, et al. A multiplex blood-based assay targeting DNA methylation in PBMCs enables early detection of breast cancer. Nat Commun. 2023;14:4724.
63. Kresovich JK, et al. Epigenetic mortality predictors and incidence of breast cancer. Aging. 2019;11:11975–87.
64. Kresovich JK, et al. Methylation-based biological age and breast cancer risk. JNCI J Natl Cancer Inst. 2019;111:1051–8.
65. Riboli E. The european prospective investigation into cancer and nutrition (EPIC): plans and progress. J Nutr. 2001;131:170S-175S.
66. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: an expanding universe. Cell. 2023;186:243–78.
67. Campisi J, et al. From discoveries in ageing research to therapeutics for healthy ageing. Nature. 2019;571:183–92.

68. Guo X, Teschendorff AE. Epigenetic clocks and inflammaging: pitfalls caused by ignoring cell-type heterogeneity. Geroscience. 2025;47:2707–19.
69. Li J, et al. Determining a multimodal aging clock in a cohort of Chinese women. Med. 2023;4:825-848.e13.
70. Reynolds LM, et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. Nat Commun. 2014;5:5366.
71. Wang Y, et al. Epigenetic influences on aging: a longitudinal genome-wide methylation study in old Swedish twins. Epigenetics. 2018;13:975–87.
72. Yuan T, et al. An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. PLoS Genet. 2015;11:e1004996-.
73. Lu T, et al. REST and stress resistance in ageing and Alzheimer's disease. Nature. 2014;507:448–54.
74. Tan Q, et al. Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort. Int J Epidemiol. 2016;45:1146–58.
75. Johansson Å, Enroth S, Gyllensten U. Continuous aging of the human DNA methylome throughout the human lifespan. PLoS ONE. 2013;8:e67378.
76. Ballas N, Grunseich C, Lu DD, Speh JC, Mandel G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. Cell. 2005;121:645–57.
77. dos Santos GA, Chatsirisupachai K, Avelar RA, de Magalhães JP. Transcriptomic analysis reveals a tissue-specific loss of identity during ageing and cancer. BMC Genomics. 2023;24:644.
78. Martinez-Jimenez CP, et al. Aging increases cell-to-cell transcriptional variability upon immune stimulation. Science. 2017;1979(355):1433–6.
79. Mansell G, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. BMC Genomics. 2019;20:366.
80. Mogilenko DA, Shchukina I, Artyomov MN. Immune ageing at single-cell resolution. Nat Rev Immunol. 2022;22:484–98.
81. Jansen RJ, et al. The effect of age on DNA methylation in whole blood among Bangladeshi men and women. BMC Genomics. 2019;20:704.
82. Zhang Q, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med. 2019;11:54.
83. Field AE, et al. DNA methylation clocks in aging: categories, causes, and consequences. Mol Cell. 2018;71:882–95.
84. Aberg KA, et al. Testing two models describing how methylome-wide studies in blood are informative for psychiatric conditions. Epigenomics. 2013;5:367–77.
85. Starnawska A, Tan Q, McGue M, Mors O, Børglum AD, Christensen K, Nyegaard M, Christiansen L. Epigenome-Wide Association Study of Cognitive Functioning in Middle-Aged Monozygotic Twins. Front Aging Neurosci. 2017;9:413. https://www.frontiersin.org/journals/aging-neuroscience/articles/10.3389/fnagi.2017.00413.
86. Awasthi N, Liongue C, Ward AC. STAT proteins: a kaleidoscope of canonical and non-canonical functions in immunity and cancer. J Hematol Oncol. 2021;14:198.
87. Angelats L, et al. Linking tumor immune infiltration to enhanced longevity in recurrence-free breast cancer. ESMO Open. 2025;10:104109.
88. Yang L, et al. Methylation of a CGATA element inhibits binding and regulation by GATA-1. Nat Commun. 2020;11:2560.
89. Ma H-Y, et al. NFIX suppresses breast cancer cell proliferation by delaying mitosis through downregulation of CDK1 expression. Cell Death Discov. 2025;11:77.
90. Han L, et al. Changes in DNA methylation from pre- to post-adolescence are associated with pubertal exposures. Clin Epigenetics. 2019;11:176.
91. Yusipov I, et al. Age-related DNA methylation changes are sex-specific: a comprehensive assessment. Aging. 2020;12:24057–80.
92. Bollepalli S, Korhonen T, Kaprio J, Anders S, Ollikainen M. Epismoker: a robust classifier to determine smoking status from DNA methylation data. Epigenomics. 2019;11:1469–86.
93. Chybowska AD, et al. A blood- and brain-based EWAS of smoking. Nat Commun. 2025;16:3210.
94. Martino DJ, 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;6:1085–94.
95. Wikenius E, Moe V, Smith L, Heiervang ER, Berglund A. DNA methylation changes in infants between 6 and 52 weeks. Sci Rep. 2019;9:17587.
96. Jacques M, et al. DNA Methylation Ageing Atlas Across 17 Human Tissues. 2025. https://doi.org/10.1101/2025.07.21.665830.
97. Moqri M, et al. Integrative epigenetics and transcriptomics identify aging genes in human blood. bioRxiv. 2024. https://doi.org/10.1101/2024.05.30.596713.
98. Palli D, et al. A molecular epidemiology project on diet and cancer: the Epic-Italy prospective study. Design and baseline characteristics of participants. Tumori J. 2003;89:586–93.
99. Zhou W, Triche JTJ, Laird PW, Shen H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. Nucleic Acids Res. 2018;46:e123–e123.
100. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res. 2017;45:e22–e22.
101. Wood SN. Generalized Additive Models: An Introduction with R. 2017.
102. Consortium, T. G. O., et al. The gene ontology knowledgebase in 2023. Genetics. 2023;224:iyad031.
103. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
104. Gu Z, Hübschmann D. Simplifyenrichment: a bioconductor package for clustering and visualizing functional enrichment results. Genomics Proteomics Bioinformatics. 2023;21:190–202.
105. Xiong Z, et al. Ewas open platform: integrated data, knowledge and toolkit for epigenome-wide association study. Nucleic Acids Res. 2022;50:D1004–9.

106. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576–89.
107. Wielscher M, et al. DNA methylation signature of chronic low-grade inflammation and its role in cardio-respiratory diseases. Nat Commun. 2022;13:2408.
108. Grolaux R. SNITCH: Semi-supervised Non-linear Identification and Trajectory Clustering for High-dimensional Data. 2026. Preprint at https://github.com/fishrscale/SNITCH.
109. Grolaux R. SNITCH: Semi-supervised Non-linear Identification and Trajectory Clustering for High-dimensional Data. 2026. Preprint at. https://doi.org/10.5281/zenodo.18203539.

Publisher's Note